

# **Určenie modelu komplexnej siete**

## **Determination of Complex Network Model**

## Zadání bakalářské práce

Student: **Iveta Kureková**  
Studijní program: B2647 Informační a komunikační technologie  
Studijní obor: 2612R025 Informatika a výpočetní technika  
Téma: **Určení modelu komplexní sítě**  
**Determination of Complex Network Model**

### Zásady pro vypracování:

Zkoumání struktury a procesů v komplexních sítích reprezentujících reálné sítě různých typů (technické, informační, sociální apod.) je v současnosti dynamicky se vyvíjející oblastí. Určování modelu sítě je jedním z prvních kroků, které se při analýze komplexní sítě zpravidla dělají. Cílem práce je implementace algoritmů pro určení modelu (typu) komplexní sítě zejména na základě distribuce stupňů vrcholů.

1. Seznamte se s komplexními sítěmi a jejich základními modely.
2. Seznamte se se základními vlastnostmi sítí, na jejichž základě se dá určit model (typ) sítě.
3. Vyberte a naimplementujte algoritmy pro určení modelu sítě.
4. Nad vybranými kolekcemi dat proveďte experimenty a jejich výsledky vhodně reprezentujte.

### Seznam doporučené odborné literatury:

1. M. E. J. Newman, Networks: An Introduction, Oxford University Press (2010), ISBN-10: 0199206651.
2. Dle pokynů vedoucího bakalářské práce.

Formální náležitosti a rozsah bakalářské práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí bakalářské práce: **RNDr. Eliška Ochodková, Ph.D.**

Datum zadání: 18.11.2011

Datum odevzdání: 07.05.2013

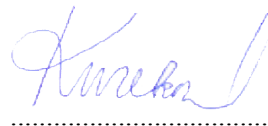


doc. Dr. Ing. Eduard Sojka  
vedoucí katedry

prof. RNDr. Václav Snášel, CSc.  
děkan fakulty


Súhlasím so zverejnením tejto bakalárskej práce podľa požiadavkov čl. 26, odst. 9 *Študijného a skúšobného rádu pre štúdium v bakalárskych programoch VŠB-TU Ostrava*.

V Ostrave 16. apríla 2013

  
.....

Prehlasujem, že som túto bakalársku prácu vypracovala samostatne. Uviedla som všetky literárne pramene a publikácie, z ktorých som čerpala.

V Ostrave 16. apríla 2013

  
.....

Chcela by som sa poďakovať vedúcej bakalárskej práce RNDr. Eliške Ochodkovej, Ph.D., za čas, ktorý mi venovala pri zodpovedaní odborných i praktických otázok, za jej odborné vedenie, poskytnuté rady, pripomienky a trpezlivosť.

## Abstrakt

Teoretická časť bakalárskej práce popisuje komplexné siete, ich vlastnosti, hlavne priemernú vzdialenosť siete, zhlukovací koeficient a distribúciu stupňov. Zahŕňa základné modely, snažiace sa vierohodne zachytiť štruktúru reálnych sietí. Prezентuje algoritmy na zistenie modelu sietí. Experimentálna časť obsahuje implementáciu aplikácie, zisťujúca exponent konektivity uzlu troma rôznymi metódami: maximálnym vierohodným odhadom, logaritmickým binningom a kumulatívnou funkciou. Za pomoci tejto aplikácie boli urobené experimenty na rôznych dátových kolekciiach s cieľom určiť model siete a exponent konektivity uzlu. Výsledky sú reprezentované grafom.

**Kľúčová slova:** komplexné siete, priemerná vzdialenosť siete, zhlukovací koeficient, distribúcia stupňov, Erdős–Rényi model, Watts-Strogats model, bezškálové siete

## Abstract

Theoretical part of the Bachelor thesis describes complex networks and their properties, mainly Average Path Length of network, Clustering Coefficient and Degree Distribution. Includes basic models, aiming to authentically capture the structure of real networks. It presents algorithms for detection of networks model. Experimental part consists of application implementation, which gathers degree exponent by three different methods: Maximum Likelihood Estimation, Logarithmic Binning and Cumulative Function. Experiments on different data collections for purpose of specification network model and degree exponent were made with the help of this application. Results are presented in plot.

**Keywords:** Complex Networks, Average Path Length, Clustering Coefficient, Degree Distribution, Erdős–Rényi Model, Watts-Strogats Model, Scale-Free Networks

## **Seznam použitých zkratk a symbolů**

ASN	– Autonomous System Number
MLE	– Maximum Likelihood Estimation

## Obsah

<b>1</b>	<b>Úvod</b>	<b>6</b>
<b>2</b>	<b>Komplexné siete</b>	<b>7</b>
2.1	Vývoj modelov komplexných sietí . . . . .	7
2.2	Základné pojmy z teórie grafov . . . . .	8
<b>3</b>	<b>Vlastnosti siete</b>	<b>10</b>
3.1	Priemerná vzdialenosť . . . . .	10
3.2	Zhlukovací koeficient . . . . .	10
3.3	Distribúcia stupňov . . . . .	11
<b>4</b>	<b>Modely siete</b>	<b>14</b>
4.1	Náhodný model $G(n,m)$ . . . . .	14
4.2	Náhodný model $G(n,p)$ . . . . .	14
4.3	Watts - Strogats model . . . . .	17
4.4	Bezškálové siete . . . . .	20
<b>5</b>	<b>Algoritmy na zistenie modelu siete</b>	<b>24</b>
5.1	Logaritmický binning . . . . .	24
5.2	Kumulatívna funkcia . . . . .	26
5.3	Maximálny vierohodný odhad . . . . .	26
<b>6</b>	<b>Experimentálna časť</b>	<b>27</b>
6.1	Analýza požiadavkov . . . . .	27
6.2	Návrh . . . . .	28
6.3	Implementácia . . . . .	29
6.4	Experimenty nad vybranými dátovými kolekciami . . . . .	33
<b>7</b>	<b>Záver</b>	<b>44</b>
<b>8</b>	<b>Literatura</b>	<b>45</b>
	<b>Přílohy</b>	<b>46</b>
<b>A</b>	<b>Príloha</b>	<b>48</b>

## Seznam tabulek

1	Tabuľka exponentov $\alpha$ z kolekcií dát . . . . .	33
---	--	----



## Seznam obrázků

1	Distribučné rozdelenie stupňov .....	12
2	Porovnanie poissonového a mocninového rozdelenia .....	13
3	Náhodný graf.....	15
4	Pravidelná mriežka, sieť malého sveta a náhodný graf .....	18
5	Postupná transformácia na náhodný graf .....	19
6	Bezškálová sieť .....	22
7	Histogram stupňovej distribúcie v lineárnom zobrazení .....	24
8	Histogram stupňovej distribúcie v logaritmickom zobrazení .....	25
9	Histogram stupňovej distribúcie v logaritmickom zobrazení .....	25
10	Užívateľské rozhranie .....	29
11	Wikipedia češtiny v lineárnom zobrazení .....	33
12	Wikipedia češtiny v logaritmickom zobrazení .....	33
13	Wikipedia češtiny v kumulatívnej funkcii .....	33
14	Wikipedia češtiny v lineárnom zobrazení .....	34
15	Wikipedia češtiny v logaritmickom zobrazení .....	34
16	Wikipedia češtiny v kumulatívnej funkcii .....	34
17	Wikipedia taliančiny v lineárnom zobrazení .....	36
18	Wikipedia taliančiny v logaritmickom zobrazení .....	36
19	Wikipedia taliančiny v kumulatívnej funkcii .....	36
20	Wikipedia taliančiny v lineárnom zobrazení .....	37
21	Wikipedia taliančiny v logaritmickom zobrazení .....	37
22	Wikipedia taliančiny v kumulatívnej funkcii .....	37
23	Wikipedia francúzštiny v lineárnom zobrazení .....	39
24	Wikipedia francúzštiny v logaritmickom zobrazení .....	39
25	Wikipedia francúzštiny v kumulatívnej funkcii .....	39
26	Wikipedia francúzštiny v lineárnom zobrazení .....	40
27	Wikipedia francúzštiny v logaritmickom zobrazení .....	40
28	Wikipedia francúzštiny v kumulatívnej funkcii .....	40
29	Spolupracujúci autori v lineárnom zobrazení .....	42
30	Spolupracujúci autori v logaritmickom zobrazení .....	42
31	Spolupracujúci autori v kumulatívnej funkcii .....	42
32	Autonómne systémy v lineárnom zobrazení .....	44

33	Autonómne systémy v logaritmickej zobrazení .....	44
34	Autonómne systémy v kumulatívnej funkcii .....	44

## Seznam výpisů zdrojového kódu

1	Metóda ReverseString vracajúca opak reťazca . . . . .	29
2	Metóda SplitWords vracajúca hodnotu v riadku . . . . .	30
3	Metóda DegreeOfEdges na výpočet stupňovej distribúcie . . . . .	30
4	Metóda WriterDegreesQuantity zapisuje do súboru . . . . .	31
5	Metóda CalculateDirection na výpočet exponentu . . . . .	31

## 1 Úvod

Cieľom bakalárskej práce je poskytnúť súhrnný pohľad na komplexné siete a spoločne s vlastnou aplikáciou skúmať štruktúru vybraných sietí. V prvej kapitole sa oboznámime, do akých oblastí siahajú komplexné siete. Ďalšia kapitola sa skladá z dvoch častí. Prvá sa zaoberá vývojom výskumu komplexných sietí a druhá zahŕňa základné pojmy z oblasti teórie grafov. V tretej kapitole sa budeme venovať vlastnostiam sietí ako priemernej vzdialenosti uzla, definujeme zhukovací koeficient a distribúciu stupňov. Štvrtá kapitola popisuje modely sietí, snažiace sa analyzovať reálne siete a v poslednej kapitole teoretickej časti sa prezentujú algoritmy na určenie modelov sietí. Hlavným cieľom experimentálnej časti je prevedenie a vyhodnotenie experimentov nad vybranými dátami pomocou naimplementovanej aplikácie. Zistíme podobnosť systémov, od seba navzájom odlišných, od sietí odkazov na články vo Wikipediách, cez autonómne systémy, až po autorov publikujúcich spoločne v oblasti IT.

## 2 Komplexné siete

Výskumy komplexných sietí siahajú do veľkého množstva oblastí. Poznatky, ktoré pri-nášajú spoločne s matematickými, štatistickými a inými vedeckými disciplínami nám dávajú lepšiu predstavu o zákonoch prírody, o mechanizmoch platiacich pre ľudstvo a jeho výtvary. Pri pozorovaní rôznych reálnych sietí zisťujeme ich podobnosť. Každá má svoje špecifiká, ale podobnosť je zrejmá. Štúdium štruktúry a vlastností rozsiahlych sietí je na začiatku vývoja pri predstave, kam všade siaha tento odbor. Postupne sa vnára všade tam, kde je pozorovateľná štruktúra siete, tam, kde chceme získať poznatky z veľkého objemu dát. Za pomoci technických vymožeností sa táto veda uplatňuje v informačných technológiách, biológii, sociálnych väzbách medzi živočíchmi, biochémií, potravinových reťazcoch, evolúcii, v odboroch venujúcich sa metabolickými premenami interakciou bielkovín a následne vytvárajúcou sa genetickou preddispozíciou. Spojením poznatkov z viacerých odborov spoločne s komplexnými sieťami otvára bránu väčším možnostiam na pochopenie zatiaľ nepoznaných zákonov vesmíru.

Príkladom siete môže byť komplexná sieť vytváraná v mozgu. "Pomocou funkčnej magnetickej rezonancie dokážeme merať nervovú aktivitu mozgu pomocou toku a okysličovania krvi. Táto intenzita je meraná pre oblasti nazývané voxely o veľkosti niekoľkých milimetrov. Skenovaním mozgu niekoľko krát nám približuje 3D obraz voxelov, ktoré sú aktívne. Týmto spôsobom zaznamenávame aktívne oblasti počas rôznych kognitívnych úloh a dávajúc nám lepšiu predstavu o mape mozgu. V tomto prípade sú voxely tvorené prepojenými neurónmi, ktoré popisujú fyzickú topológiu sieťovej štruktúry. Počas riešení úloh sa prepojenia menia, avšak ak sú dva voxely v korelácii, hovoríme o funkčnom prepojení a vznikajú funkčné siete mozgu. Funkčné siete sa odlišujú v závislosti od riešených úloh. Zistilo sa, že tieto siete majú charakter sietí malého sveta a bezškálových sietí [1]." Viac informácií ohľadom topológie funkčných sietí mozgu nájdeme v práci Petra Náthera a Márie Markošovej [2].

### 2.1 Vývoj modelov komplexných sietí

V spoločnej práci "On random graphs" maďarskí matematici Paul Erdősa a Alfréd Rényi v roku 1959 definovali náhodný model ako graf vznikajúci náhodným procesom. Postupne sa časom zistilo, že tento model neodzrkadľoval vlastnosti a štruktúru komplexných sietí. Neskôr v roku 1998, prišli na rad s novou myšlienkou Watts a Strogatz o teórii vzniku

malého sveta, približujúc sa k skutočnému obrazu reálnych sietí v prírode či technike. Výmena informácií výskumných firiem v rámci modelov je predmetom ďalších štúdií. Prehľad štúdií malého sveta môžeme nájsť v prácach Jona Kleinberga z roku 1999 [3].

Bezškálové siete najvierohodnejšie mapujú komplexné siete. Medzistupňami medzi týmito modelmi bolo viac, avšak v práci sa budeme venovať tými najznámejšími, vyššie uvedenými. Každý komplexný systém sa dá popísať ako graf, ktorý popisuje vzťahy a interakcie medzi jednotlivými časťami systému. Analýzou získavame poznatky o vlastnostiach a štruktúre siete, ktoré študujeme a to sú: priemerná vzdialenosť, zhukovací koeficient a distribúcia stupňov.

## 2.2 Základné pojmy z teórie grafov

Teória grafov tvorí základ našich úvah o sieťach. Pár storočí po Leonhardovi Eulerovi a jeho predložení matematického dôkazu týkajúceho sa kráľoveckého hlavolamu, sa z nej stala vyspelá disciplína, ku ktorej prispela väčšina veľkých matematikov ako Augustin-Louis Cauchy, Wiliam Hamilton, Arthur Cayley, Gustav Kirchhoff a George Pólya. Informácie o probléme vzniknutom v Kráľovci môžeme nájsť v tejto publikácii [4].

### 2.2.1 Graf

Chcem upozorniť na fakt, že v literatúre môžeme nájsť odlišnú definíciu grafu, napr. v publikácii od Jiřího Demela [5].

Grafom nazveme usporiadanú dvojicu  $G = (V, H)$ , kde  $V$  je neprázdna konečná množina a  $H$  je množina neusporiadaných dvojíc typu  $u, v$  takých, že  $u \in V, v \in V$  a t.j.  $H \subseteq u, v | u, v \in V \subset V \circ V$ . Prvky množiny  $V$  nazývame vrcholmi a prvky množiny  $H$  hranami grafu  $G$  [6].

### 2.2.2 Digraf

Digrafom (orientovaným grafom) nazveme usporiadanú dvojicu  $\vec{G} = (V, H)$ , kde  $V$  je neprázdna konečná množina a  $H$  je množina usporiadaných dvojíc typu  $(u, v)$  takých, že  $u \in V, v \in V$  t.j.  $H \subseteq u, v | u, v \in V \subset V \times V$ . Prvky množiny  $V$  nazývame vrcholmi a prvky množiny  $H$  orientovanými hranami digrafu  $\vec{G}$ . Nech  $\vec{G} = (V, H)$  je digraf,

$u \in V, v \in V, h \in H$ . Hovoríme, že orientovaná hrana  $h$  vychádza z vrchola  $u$ , alebo že vrchol  $u$  je začiatkový vrchol orientovanej hrany  $h$ , ak  $h = (u, x)$  pre niektoré  $x \in V$ . Hovoríme, že orientovaná hrana  $h$  vchádza do vrchola  $v$ , alebo že vrchol  $v$  je koncový vrchol orientovanej hrany  $h$ , ak  $h = (y, v)$  pre niektoré  $y \in V$ . Orientovaná hrana  $h$  je incidentná s vrcholom  $v$ , ak hrana  $h$  vchádza do vrchola  $v$  alebo vychádza z vrchola  $v$ .  $H^+(v)$  je množina všetkých orientovaných hrán digrafu  $\vec{G}$  vychádzajúcich z vrchola  $v$ .  $H^-(v)$  je množina všetkých orientovaných hrán digrafu  $\vec{G}$  vchádzajúcich do vrchola  $v$ .  $V^+(v)$  je množina koncových vrcholov všetkých hrán z  $H^+(v)$  a  $V^-(v)$  je množina začiatkových vrcholov všetkých hrán z  $H^-(v)$  [6].

### 2.2.3 Diagram grafu

Graf reprezentujeme graficky a príslušný obrázok voláme diagram grafu. Diagram grafu  $G = (V, H)$  v priestore  $P$  je množina  $B$  bodov a množina  $S$  súvislých čiar v priestore  $P$  takých, že

- každému vrcholu  $v \in V$  zodpovedá práve jeden bod  $b_v \in B$  a každému bodu  $b \in B$  zodpovedá práve jeden vrchol  $v \in V$  (t. j.  $b_v = b$ ), pričom pre  $u, v \in V, u \neq v$  je  $b_u \neq b_v$ .
- každej hrane  $h \in H$  zodpovedá práve jedna čiara  $s_h \in S$  a každej čiare  $s \in S$  zodpovedá práve jedna hrana  $h \in H$  (t. j.  $s = s_h$ ), pričom pre  $h, k \in H, h \neq k$  je  $s_h \neq s_k$ .

Ak  $h = (u, v) \in H$ , potom čiara  $s_h$  má koncové body  $b_u, b_v$  [6].

### 2.2.4 Ostatné varianty grafov

Variety grafov ako pseudografy, obsahujúce smyčky alebo multigrafy s viacnásobnými hranami, či migrafy, obsahujúce orientované i neorientované hrany sa môžu hodiť v určitých prípadoch, avšak v práci ich využívať nebudeme, pretože pre určenie modelu komplexnej siete si vystačíme s prostým grafom, čiže grafom, v ktorom násobnosť každej hrany je rovná najvyššiemu.

## 3 Vlastnosti siete

### 3.1 Priemerná vzdialenosť

Medzi základné vlastnosti konceptu topológie siete patrí priemerná vzdialenosť siete. Definujeme ju ako priemerný počet krokov vykonaných po najkratších cestách pre všetky možné páry vrcholov. Nadobudne teda takú hodnotu najkratšej cesty, ktorá bude najviac pravdepodobná na spojenie jedného vrcholu s druhým. Príkladom tejto vzdialenosti môže byť počet kliknutí, ktoré povedú z jednej stránky na druhú alebo počet ľudí potrebných na spojenie s nami a s hociktorým človekom na zemeguli. Majme graf  $G(V)$  s vrcholmi  $V$ . Najkratšiu vzdialenosť medzi  $v_1$  a  $v_2$  označme  $d(v_1, v_2)$ , kde  $v_1, v_2 \in V$ . Súdime, že ak  $v_1 = v_2$  alebo je  $v_2$  nedosiahnuteľné z  $v_1$ , potom platí  $d(v_1, v_2) = 0$ . Z toho dokážeme vyvodit' všeobecnú priemernú vzdialenosť  $L_G = \frac{1}{(n(n-1))} \sum_{(i,j)} d(v_i, v_j)$ , kde  $n$  je počet vrcholov v  $G(V)$  [7].

Zisťovať a efektívne využívať priemerné vzdialenosti v reálnych sieťach je veľmi výhodné, šetríme časom a technickými prostriedkami a v neposlednej rade sa to pozitívne premietne do životného prostredia, ktoré sa postupne stáva prioritou v nasledujúcich rokoch a desaťročiach v technologickej sfére. Ako príklad uvediem sieť World Wide Web, kedy minimalizované priemerné vzdialenosti uľahčujú a zrýchľujú prenos informácií a znižujú náklady na energetické straty.

### 3.2 Zhukovací koeficient

Ďalšou meranou charakteristickou črtou, ktorá vystihuje štruktúru grafu je zhukovací koeficient. Táto veličina priepustnosti siete opisuje, do akej miery vrcholy inklinujú k vytváraniu skupín. Je definovaná pravdepodobnosťou, že dva vrcholy, ktoré sú susedmi s tým istým uzlom budú aj navzájom susedmi. Koeficient, na základe ktorého meriame celú sieť nazývame globálny koeficient, kým koeficient lokálny mapuje zhuk samostatne stojacich vrcholov.

#### 3.2.1 Globálny zhukovací koeficient

Je založený na trojiciach vrcholov, kde je táto trojica vrcholov spojená buď dvoma hranami, inými slovami tvorí otvorený triplet alebo tromi hranami (uzavretý triplet, čiže úplný



graf o troch vrchoch). Globálny koeficient vyčíslime pomerom počtu uzavretých trojíc, inými slovami trojnásobným počtom trojuholníkov (t. z. trojuholník = uzavretá trojica) a celkovým počtom trojíc, uzavretých aj otvorených:  $C = \frac{(3*pt)}{\binom{t}{3}} = \frac{(ut)}{\binom{t}{3}}$ , kde  $pt$  je počet trojuholníkov,  $t$  celkový počet pripojených trojíc a  $ut$  predstavuje počet uzavretých trojíc [7].

### 3.2.2 Lokálny zhlukovací koeficient

Majme graf  $G = (V, E)$ , kde  $V$  predstavujú vrcholy a  $E$  hrany medzi nimi. Potom hrana  $e_{i,j}$  spája vrchol  $v_i$  s vrcholom  $v_j$ . Potom susedstvo  $N_i$  pre vrchol  $v_i$ , je definované  $N_i = \{v_j : e(i, j) \in E \wedge e(j, i) \in E\}$ . Majme  $k_i$  ako počet susediacich vrcholov s  $v_i$ . Potom sa zhlukovací koeficient  $C(i)$  vrcholu  $i$  definuje ako pomer počtu existujúcich hrán medzi susedmi vrcholu  $i$  a počtu všetkých možných hrán medzi nimi.

Pre orientovaný graf je hrana  $e_{i,j}$  a  $e_{j,i}$  rozdielna, preto pre každého suseda  $N_i$  je možných  $(k_i)(k_i - 1)$  hrán, ktoré by mohli existovať medzi susediacimi vrcholmi. Potom je formulácia lokálneho zhlukovacieho koeficientu pre orientovaný graf nasledovná:  $C_i = \frac{(|\{e_{j,k}: v_j, v_k \in N_i, e_{j,k} \in E\}|)}{(k_i(k_i - 1))}$ . V neorientovanom grafe sú hodnoty hrán  $e_{i,j}$  a  $e_{j,i}$  ekvivalentné, preto platí, ak vrchol  $v_i$  má  $k_i$  susedov, potom možných hrán medzi susednými vrcholmi je  $\frac{(k_i(k_i - 1))}{2}$ . Potom lokálny zhlukovací koeficient pre neorientovaný graf znie  $C_i = \frac{(2|\{e_{j,k}: v_j, v_k \in N_i, e_{j,k} \in E\}|)}{(k_i(k_i - 1))}$  [9].

## 3.3 Distribúcia stupňov

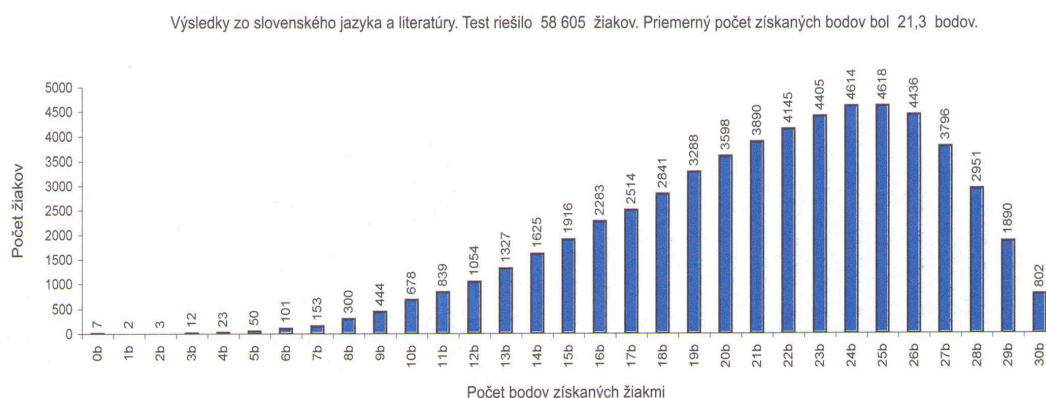
### 3.3.1 Poissonové rozdelenie

Siméon Denis Poisson, francúzsky fyzik a matematik aplikoval do oblasti teórie pravdepodobnosti a štatistiky nový typ pravdepodobnostného rozdelenia. Je to jeden z diskrétnych rozdelení pravdepodobností. Kalkuluje pravdepodobnosť výskytu udalostí pre danú priemernú hodnotu  $\lambda$ .  $\lambda = \frac{k}{n}$ , kde  $k$  je množstvo udalostí, ktoré nastali a  $n$  počet jednotiek (napr. časových, kedy jav nastal). Toto rozdelenie sa používa ako aproximácia binomického rozdelenia pre veľký počet jednotiek (napr. časových), pre  $n \rightarrow \infty$  a malú pravdepodobnosť výskytu sledovaného javu v jednom pokuse, teda pre  $p \rightarrow 0$ . Pravdepodobnostnú funkciu poissonového rozdelenia s parametrom  $\lambda$  má nespojitá náhodná veličina  $X$ , ktorej tvar zapíšeme ako  $P(X = x) = e^{(-\lambda)} \frac{\lambda^x}{x!}$  pre  $\lambda > 0$ .

Predstavme si, že priemerný počet kazov v optickom kábli o dĺžke 50 m je 1,2. Aká bude pravdepodobnosť, že budú práve tri kazy o dĺžke 150 m? Priemerný počet kazov v 150 m je 3,6. Potom pravdepodobnosť, že v kábli budú tri kazy je  $\frac{e^{-3,6}(3,6)^3}{3!} = 0,212$ , t. j. 21,2 % [11].

Poissonové rozdelenie má výrazné maximum, takže väčšina uzlov má práve priemerný stupeň, t.j. počet hrán. Po oboch stranách rýchlo klesá, odchýlky sú vzácne. Toto rozdelenie predpovedá pri 7 miliardách ľudí to, že každý z nás má priemerný počet priateľov a známych. Pravdepodobnosť, že niekto má extrémne veľa, či málo priateľov je exponenciálne malá. Poissonovo rozdelenie má rôznorodé využitie, v neposlednom rade modeloch náhodných grafov. Teória grafov predpovedá, že keď rozdelíme sociálne väzby náhodne, dostaneme demokratickú spoločnosť, kde sú všetci priemernými a len málokto sa odchyľuje od normy. Získame sieť, kde je normálne byť priemerný. V modeli Erdősa a Réneyiho sú najrozšírenejšie priemerné hodnoty [4].

Obrázok 1 zobrazuje štatistiku vydanú Ministerstvom školstva z celoslovenského testovania žiakov 9. ročníkov základných škôl, tzv. Monitor 9, zo šk. roku 2004/2005 z predmetu Slovenský jazyk a literatúra. Počet žiakov, ktorí boli testovaní, nadobudol hodnotu 58 605 žiakov. Vidíme, že rozdelenie počtu bodov na žiakov sa približuje poissonovému distribučnému rozdeleniu, nie mocninovému zákonu.

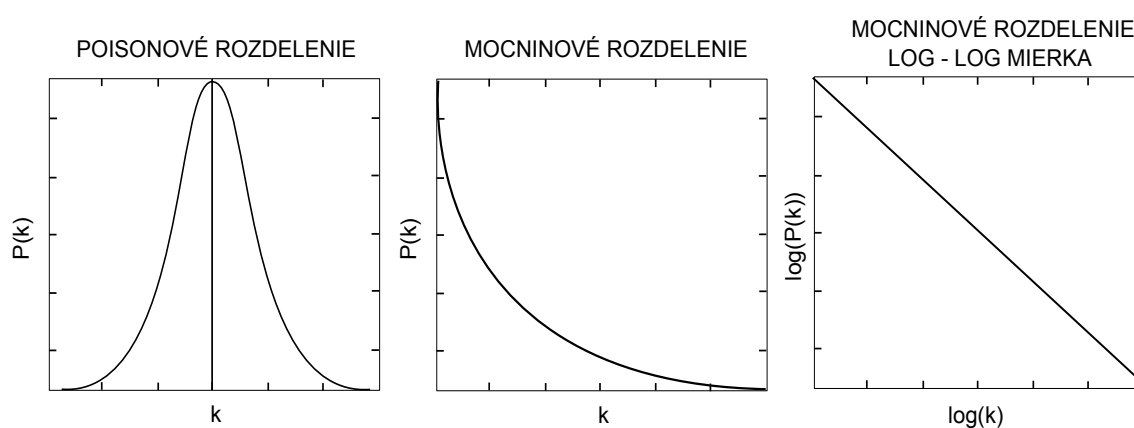


Obrázok 1: Distribučné rozdelenie stupňov

### 3.3.2 Mocninové rozdelenie

Mocninová závislosť je špeciálnym prípadom polynomickej funkcie jednej premennej, v ktorej vystupuje závislá premenná  $x$  obsahujúca exponent  $k$ . Jej podoba znie:  $f(x) = ax^k + o(x^k)$ , kde  $a$  a  $k$  sú konštanty a  $o(x^k)$  je vzhľadom k  $ax^k$  asymptoticky menšia funkcia. Pre mocninovú funkciu platí  $f(cx) \propto f(x)$ , kde je  $c$  konštanta a znamená, že zväčšením argumentu konštantným pomerom sa zmení mierka funkcie, ale nie jej tvar. Pre znázornenie mocninového zákona sa používa tvar  $\log f(x) = \log b + \alpha \log x$ . Parameter  $\alpha$  predstavuje sklon tejto lineárnej závislosti, kde ho konštanta argumentu  $b$  neovplyvňuje [12]. Mocninový zákon sa uplatňuje v reálnych sieťach, ktoré patria medzi bezškálové siete, ako napríklad veľkosť miest a firiem, makroekonomické ukazovatele, frekvencia slov vo vetných skladbách, rozdelenie príjmov a bohatstva.

Na obrázku 2 v strede vidíme, že histogram takejto siete, riadiaca sa mocninovým zákonom sa líši od zvonovitej krivky poissonového rozdelenia (vľavo). Nemá výrazné maximum a namiesto toho má sieť veľké množstvo malých udalostí a málo veľkých udalostí. Na obrázku 2 vpravo, vidíme preloženú krivku mocninového rozdelenia do logaritmicko-logaritmicko mierky. Samozrejme, že nie všetky siete podliehajú mocninovému zákonu. Keby sa napr. výška ľudí riadila mocninovým zákonom, na planéte by sme boli poväčšine nízki škriatkovia a výnimočne by sme mohli stretnúť aj 50-metrového človeka [4].



Obrázok 2: Porovnanie poissonového a mocninového rozdelenia

## 4 Modely siete

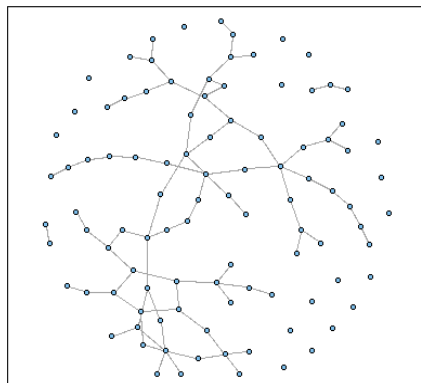
### 4.1 Náhodný model $G(n, m)$

Náhodným pospájaním vrcholov  $n$  pomocou hrán  $m$ , vylučujúc násobné hrany a slučky, vzniká náhodný graf, t. j. graf, kde náhoda určuje, ktoré páry vrcholov budú spojené hranou. Náhodný model  $G(n, m)$  je definovaný pravdepodobnosťou rozdelenia  $P(G)$  hrán medzi vrcholmi nad grafmi s daným počtom vrcholov  $n$  a hrán  $m$ , ktoré môžu existovať. Platí  $P(G) = \frac{1}{\Omega}$  pre možný graf, kde  $\Omega$  je počet jednotlivých možných grafov. Ako príklad uvediem graf  $G = (3, 2)$ , kde počet možných grafov, ktoré sa dajú zostrojiť s tromi vrcholmi a dvoma hranami sú tri. Potom pravdepodobnosť, že bude vybraný graf je  $\frac{1}{3}$ . V analytických meraniach je okrem vo vyššie uvedenej kapitole užitočnou vlastnosťou aj priemer siete  $l(G)$ , definovaná ako maximálna vzdialenosť medzi dvoma ľubovoľnými vrcholmi v sieti:  $\langle l \rangle = \sum_{(G)} P(G)l(G) = \frac{1}{\Omega} \sum_{(G)} l(G)$ . Typy grafov  $G(n, m)$  s veľkým alebo malým priemerom siete neodzrkadľujú typické správanie  $G(n, m)$ . Niektoré vlastnosti náhodného grafu  $G(n, m)$  sa dajú presne vyčísliť, príkladom môže byť výpočet priemerného stupňa  $\langle k \rangle = \frac{2m}{n}$ , kde  $m$  je priemerný počet hrán. Napr. priemerný stupeň pre  $G(5, 6) = \frac{10}{5} = 2$ . Obtiažne vypočítateľné vlastnosti kalkuluje pomocou podobného modelu  $G(n, p)$  [10].

### 4.2 Náhodný model $G(n, p)$

V predchádzajúcom modeli  $G(n, p)$  sme uvádzali celkové množstvo hrán  $m$ , v modeli  $G(n, p)$  namiesto nich uvádzame pravdepodobnosť  $p$  existujúcich hrán  $m$  medzi vrcholmi. Majme zvolený počet vrcholov  $n$  a s pravdepodobnosťou  $p$  umiestnime hranu medzi uzly. Model  $G(n, p)$  je daný súborom sietí s  $n$  vrcholmi, v ktorých každý možný graf  $G$  sa javí s pravdepodobnosťou  $P(G) = p^m (1-p)^{\binom{n}{2}-m}$  kde  $m$  je počet hrán v grafe.

$G(n, p)$  model sa pripisuje dvom maďarským matematikom Paulovi Erdősovi a Alfrédovi Rényiovi, ktorý ho publikovali na začiatku 60. rokov minulého storočia. Pripisuje sa mu viacero názvov, ako "Erdős-Rényi náhodný graf" alebo aj "Poissonový náhodný graf." Patrí medzi fundamentálny model, spomedzi všetkých typov náhodných modelov, ktoré kedy boli publikované, preto sa zaužíval i všeobecný názov "náhodný model" práve pre tento typ [10]. Na obrázku 3 [19] máme možnosť pozorovať náhodný graf, kde pravdepodobnosť, že je uzol spojený s ďalším je 2%.



Obrázok 3: Náhodný graf

#### 4.2.1 Vlastnosti náhodného modelu

**4.2.1.1 Priemerný počet hrán a priemerný stupeň vrcholu** Počet grafov so stanoveným počtom vrcholov  $n$  a hrán  $m$  je vyčísliteľný  $\binom{n}{m}$ . Majme napr. graf  $G(4, 2)$ , potom  $\binom{4}{2} = \binom{6}{2} = 15$ . Existuje celkovo 15 rôznych grafov  $G(4, 2)$ . Každý z týchto grafov má rovnakú pravdepodobnosť  $P(G)$  uvedenú v predchádzajúcej kapitole a preto celková pravdepodobnosť vybrania grafu s  $m$  hranami zo všetkých možných grafov je daná  $P(m) = \binom{n}{m} p^m (1-p)^{\binom{n}{2}-m}$ , čo je štandardné binomické rozdelenie pravdepodobnosti.

Potom priemerný počet hrán  $m$  v grafe je  $\langle m \rangle = \sum_{m=0}^{\binom{n}{2}} m P(m) = \binom{n}{2} p$ . Z toho plynie, že celkový počet hrán je rovnaký očakávanej hodnote  $p$  medzi párom vrcholov znásobený počtom všetkých párov v grafe.

Na výpočet priemerného stupňa vrcholu v náhodnom grafe  $G(n, p)$  využijeme uvádzaný výpočet priemerného stupňa s presným počtom hrán  $\langle k \rangle = \frac{2m}{n}$ , využitý v modeli  $G(n, m)$  a preto priemerný stupeň vrcholu v  $G(n, p)$  je daný  $\langle k \rangle = \sum_{m=0}^{\binom{n}{2}} \frac{2m}{n} P(m) = \frac{2}{n} \binom{n}{2} p = (n-1)p$ . Často sa v literatúre označuje priemerný stupeň vrcholu náhodného grafu písmenom  $c$  alebo  $z$ , preto  $z = (n-1)p$ . Z rovnice plynie, že očakávaný počet hrán spojených s vrcholom je rovnaký ako očakávaná pravdepodobnosť  $p$  medzi vrcholom a ďalším vrcholom znásobeným o  $n-1$  ďalších vrcholov [10].

**4.2.1.2 Priemerná vzdialenosť a súvislosť** Ak je v grafe počet hrán  $m$  nízky,  $z < 1$ , potom graf obsahuje vysoký počet súvislých komponentov. Graf je zložený z izolovaných

stromov, priemerná vzdialenosť odpovedá priemernej vzdialenosti stromu a najväčšia komponenta má počet vrcholov najviac  $O(\log n)$ . Ak je počet hrán  $m = \frac{n}{2}$ , teda  $z > 1$ , najväčšia komponenta má veľkosť  $\Theta(n)$  a druhá najväčšia  $O(\log n)$ . Ak je  $z > 3.5$ , priemerná vzdialenosť grafu je rovná priemernej vzdialenosti  $L$  najväčšej komponenty a je úmerná  $L = \frac{\log n}{\log z}$ . Ak  $z > \log n$ , graf je súvislý a priemerné vzdialenosti nadobúdajú hodnoty okolo  $L = \frac{\log n}{\log z}$ . Ak  $z = 1$ , nastane zmena, vedúca k vzniku rozsiahlej komponenty  $O(n^{\frac{2}{3}})$ , v ktorej platí mocninový zákon [12].

**4.2.1.3 Distribúcia stupňov** Vrchol v grafe je spojený s pravdepodobnosťou  $p$  k ďalšiemu vrcholu  $n - 1$ . Potom pravdepodobnosť spojenia prislúchajúcim vrcholom je daná  $p^k = p^k(1 - p)^{(n-1-k)}$ . Je  $\binom{n-1}{k}$  ciest na výber práve  $k$  ďalších vrcholov a preto celková pravdepodobnosť, že bude spojený práve  $k$  ďalším je  $p_k = \binom{n-1}{k} p^k (1 - p)^{n-1-k}$ , kde  $n$  je celkový počet vrcholov v grafe. Vidíme, že  $G(n, p)$  má binomický stupeň rozdelenia.

U vlastnostiach rozsiahlych sietí, kde je  $n$  vysoké, častokrát vypozerujeme približne nemenný priemerný stupeň aj keď sa sieť zväčšuje. Napr. počet praktických lekárov na osobu nie je závislý na zvyšujúcej sa počte osôb na zemi. Rovnica uvedená vyššie  $p = \frac{z}{n-1}$ , nám umožňuje napísať  $\ln[(1 - p)^{n-1-k}] = (n - 1 - k) \ln(1 - \frac{z}{n-1}) \simeq -(n - 1 - k) \frac{z}{n-1} \simeq -z$ , tu sme rozšírili logaritmus ako Taylorov rad, kde  $n \rightarrow \infty$ . Pre značne veľké  $n$  platí  $\binom{n-1}{k} = \frac{(n-1)!}{(n-1-k)!k!} \simeq \frac{(n-1)^k}{k!}$  a teda  $p_k = \frac{(n-1)^k}{k!} p^k e^{-z} = \frac{n-1^k}{k!} (\frac{z}{n-1})^k e^{-z} = e^{-z} \frac{z^k}{k!}$  [10]. Z rovnice vidíme, že  $G(n, p)$  má Poissonovo stupňové rozdelenie. Odtiaľto pochádza názov Poissonov náhodný graf.

**4.2.1.4 Zhukovací koeficient** Ako sme spomínali, pravdepodobnosť, že dva vrcholy sú susedmi je  $p = \frac{z}{n-1}$ , a platí,  $C = p$ , a teda  $C = \frac{z}{n-1}$  [10]. Zhukovací koeficient sa odlišuje od reálnych sietí, väčšina z nich, zhukovací koeficient má tendenciu smerovať k nule, kde  $n \rightarrow \infty$ .

Pozorovaním sa zistilo, že model  $G(n, p)$  nie je najvhodnejší pre skúmanie reálnych sietí ako štruktúrovanie telefónnych sietí, sociálnych väzbách ľudí či chemických reakcií buniek. Reálne siete sa totiž vyznačujú centrami, vysokým zhukovacím koeficientom a rozdelenie podľa mocninového zákona nemá výrazné maximum [4].

## 4.3 Watts - Strogats model

### 4.3.1 Vznik pojmu siete malého sveta

Počiatky vzniku grafov malého sveta spadajú do roku 1967, kedy sociálny psychológ Stanley Milgram z Yalskej univerzity učinil pokus, týkajúci sa ľudskej spoločnosti. Napísal množstvo listov určených pre jeho priateľa, žijúceho v Bostone a rozdelil ich náhodne vybraným osobám v Nebrasce. Každý z nich bol požiadaný, aby sa pokúsil doručiť list adresátovi tak, aby bol doručovaný od človeka k človeku, pretože sa predpokladalo, že ani jeden nebude poznať adresáta. Podmienkou bolo, že dvaja ľudia, medzi ktorými práve prebiehalo doručovanie, sa museli poznať. Po skončení pokusu Milgram zistil, že priemerný počet ľudí, potrebných na doručenie pošty svojmu priateľovi bolo šesť. Z toho vyvodil záver, že priemerná vzdialenosť medzi dvomi náhodne vybranými ľuďmi na zemeguli je šesť. Do povedomia sa tak dostal pojem šesť stupňov odlúčenia [4].

Tento experiment nevedomky naviazal na prvú myšlienku maďarského spisovateľa Frigyesa Karinthyho, ktorý v poviedke Články reťazcov (Chains) z roku 1929 odhadoval, že medzi dvoma ľubovoľnými ľuďmi sa nájde reťazec známostí nie dlhší ako päť osôb [13]. Fenomén malého sveta nečakaným spôsobom vysvetlil povahu moderného sveta ako o neriadenej obrovskej sieti vzťahov a pritom sú tieto možnosti veľmi blízke k teoretickému optimu. Ľudskú spoločnosť ako "najširší obzor komunikácie" poňal nemecký sociológ Niklas Luhmann. Tieto myšlienky spopularizoval vo filme Six Degrees of Separation americký dramatik John Guare.

Svet je "malý", preto sa sieťam, napr. aj pre skúmanie medziľudských vzťahov hovorí "siete malého sveta." Vyznačujú sa preto vysokým stupňom zhľukovania. Pozorovaním sa teda zistilo, že na rozdiel od  $G(n, p)$  v reálnych sieťach je vysoká pravdepodobnosť, že ak sa jeden pozná s druhým a tretím, druhý i tretí sa tiež poznajú [4].

### 4.3.2 Definícia Watts Strogats modelu

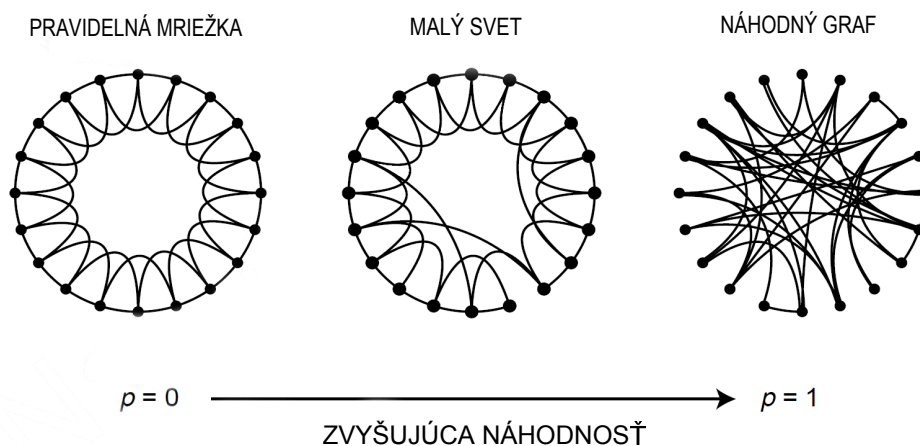
Na získanie reálnejších výsledkov ako vykazoval model  $G(n, p)$ , bol definovaný nový model. Dostal názov podľa svojich autorov a vznikol Watts-Strogats model. O rok neskôr, presnejšie v 1999 - 2000 bol publikovaný vylepšený model Watts-Newman. Východiskový graf je zostavený z vrcholov  $n$  usporiadaných do kruhu, kde každý vrchol  $n$  je spojený s

$k$  bližšími susediacimi vrcholmi a zároveň s malou nenulovou pravdepodobnosťou  $p$  aj s vrcholmi, ktoré nesusedia s vrcholom, nazývané skratky.

Do úvahy neprichádzajú násobné hrany a slučky. Vyznačuje sa krátkymi priemernými vzdialenosťami a vysokým zhukovacím koeficientom. Malá pravdepodobnosť  $p$  vytvára pravidelnú mriežku, zatiaľ čo vysoká získa vlastnosti náhodného grafu.

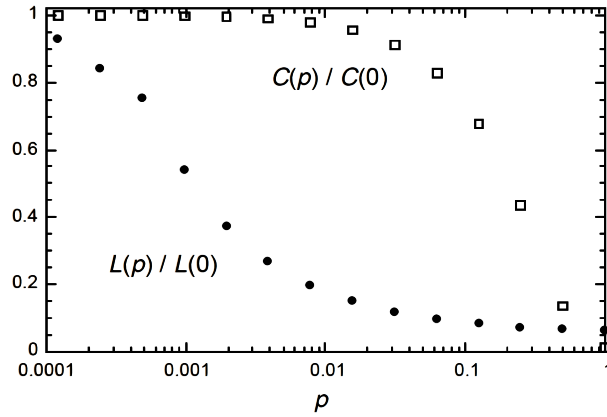
Na obrázku 4 máme znázornené 3 grafy, kde má každý rôznu pravdepodobnosť spojenia vrcholu s nesusedným vrcholom. Vľavo je graf pravidelnej mriežky, vyznačujúci sa nulovou pravdepodobnosťou, že bude spojený aj s nesusednými vrcholmi, kde pri náraste pravdepodobnosti v strednom grafe vidíme reálnu sieť Wattsovými a Strogatsovými očami - sieť malého sveta. Ďalším postupným odoberaním susedných spojení a ich následným náhodným spojením s nesusednými vrcholmi v grafe napravo máme možnosť vidieť zmenu zo siete malého sveta na náhodný graf. Postupným prerodom pravidelnej mriežky k náhodnému grafu a zmeny ich štrukturálnych vlastností môžeme vidieť na obrázku 5, ako sa mení ich priemerná vzdialenosť  $L(p)$  a zhukovací koeficient  $C(p)$  a tým aj typ modelu.

Numerická simulácia Watta a Strogatsa (obrázok 5) na zvyšovanie pravdepodobnosti  $p$  od 0 po 1 zistila, že čím  $p$  nadobúda vyššie hodnoty, tým sa rýchlejšie znižuje priemerná vzdialenosť  $L(p)$  a pomalšie klesá zhukovací koeficient  $C(p)$  [9].



Obrázok 4: Pravidelná mriežka, sieť malého sveta a náhodný graf [9]





Obrázok 5: Postupná transformácia na náhodný graf [9]

### 4.3.3 Vlastnosti pravidelnej mriežky

Pre výpočet  $k$  spojení medzi susediacimi vrcholmi vo všeobecnosti platí, že susedný vrchol  $\frac{k}{2}$  vrchola  $i$  sa môže pripojiť k  $\frac{k}{2} - 1$  susedným vrcholom, sused  $\frac{k}{2} - 1$  sa môže pripojiť k  $1 + \frac{k}{2} - 1$  susedným vrcholom, sused  $\frac{k}{2} - 2$  má  $2 + \frac{k}{2} - 1$  susedov a susedný vrchol  $i$  v počte krokov 1 má  $2(\frac{k}{2}-1)$  susedov. Sčítaním týchto spojení, podelením 2 (pre neorientovaný graf) a vynásobením 2, nakoľko vrchol  $i$  má susediace vrcholy z pravej aj ľavej strany, získame počet spojení medzi susediacimi vrcholmi  $i$ :  $\sum_{j=0}^{\frac{k}{2}-1} \frac{k}{2} + i - 1 = \frac{3}{8}k(k-2)$ .

Maximálny počet spojení  $k$  susedmi je daný  $k(\frac{k-1}{2})$  a zhukovací koeficient pre mriežku je  $C = \frac{3(k-2)}{4(k-1)}$ . Priemerná vzdialenosť, menená s lineárne rastúcim  $n$ , je daná  $L \approx \frac{n}{2k}$ . K blížiacej pravdepodobnosti  $p = 1$ , preberá vlastnosti náhodného modelu  $G(n, p)$ , má malý zhukovací koeficient vyčíslený  $C \approx \frac{k}{n}$  a nízku priemernú vzdialenosť, menenú s logaritmicky rastúcim  $n$ ,  $L \approx \frac{\ln n}{\ln k}$  [12].

### 4.3.4 Priemerná vzdialenosť malého sveta

Na výpočet priemernej vzdialenosti  $L$  platí  $L = \frac{n}{k} f(nkp)$ , kde  $f(x) = \frac{1}{2\sqrt{x^2+2x}} \tanh^{-1}(\sqrt{\frac{x}{x+2}})$  a  $nkp$  je očakávaný počet skratiek. Najlepšie výsledky získame tam, kde  $nkp$  má veľmi vysokú alebo veľmi nízku hodnotu. Ak  $nkp \gg 1$  platí  $L = \frac{\ln nkp}{k^2 p}$ . Inými slovami, ak je počet skratiek väčší ako 1, priemerná vzdialenosť  $L$  vzrastie logaritmicky s počtom vrcholov  $n$  [17].

### 4.3.5 Zhukovací koeficient malého sveta

Zo zovšeobecnenej formuly zhukovacieho koeficientu ukázali Barrat a Weigt, že platí  $C \sim \frac{3(K-1)}{2(2K-1)}(1-p)^3$  kde  $n \rightarrow \infty$  [14]. Hodnota  $C$  je značne vysoká dokonca aj pre vysoké hodnoty  $p$  a hodnota  $L$  je nízka aj pre nízke hodnoty  $p$ . Wats-Strogats model má významné rysy, kde sa udržiava vysoká hodnota pravdepodobnosti  $p$ , ktorá tvorí grafy malého sveta s významným zhukovaním a ktoré boli inšpirované mnohými výskumami, najmä Newmanom. Aj keď pridáme hranu medzi dva vzdialené vrcholy a pomocou numerických výpočtov majú Watts-Strogatsove siete vysoký koeficient zhukovania  $C$  a nízku priemernú vzdialenosť  $L$ , tieto tvrdenia nie sú považované za úplne presné a namiesto toho Watts spoločne s Newmannom definovali  $p$  ako pravdepodobnosť pridania skratky medzi každým párom vrcholov. Tento model má podobné správanie pre stredné hodnoty  $p$ , ale nie pre  $p$  blízke k hodnotám 1. Pre  $n \rightarrow \infty$ , zhukovací koeficient Newman modelu je daný  $C = \frac{3(k-1)}{2(2k-1)+4kp+4kp(p+2)}$  [17].

### 4.3.6 Distribúcia stupňov malého sveta

Stupňová distribúcia malého sveta neodzrkadľuje práve najlepšie reálne siete, ale nebolo to ani prvotným cieľom dokázať. Vo verzii modelu, kde sa hrany nepridávajú, kde má vrchol stupeň aspoň  $2k$ , kde sa tvorí pravidelná mriežka a kde sú počty skratiek distribuované binomicky, tak pravdepodobnosť  $p_j$ , že bude mať stupeň  $j$  je  $p_j = \binom{L}{j-2k} \left[\frac{2kp}{L}\right]^{j-2k} \left[1 - \frac{2kp}{L}\right]^{L-j-2k}$  pre  $k \leq j$  a  $p_j = 0$  pre  $j < 2k$ . Pre verziu modelu, kde sa hrany znovu pripoja a kde má vrchol stupeň menší než  $2k$  je pravdepodobnosť, že vrchol bude spojený hranou daná  $p_j = \sum_{n=0}^{\min(j-k, k)} \binom{k}{n} (1-p)^{k-n} \frac{(pk)^{j-k-n}}{(j-k-n)!} e^{-pk}$  pre  $j \geq k$ , a  $p_j = 0$  pre  $j < k$ . Distribúcia stupňov neodpovedá reálnym sieťam, dostávame Poissonové rozdelenie, podobne ako u náhodných modelov sietí [14].

## 4.4 Bezškálové siete

### 4.4.1 Pojem bezškálovosť

Jav vo všeobecnosti označíme ako bezškálový, kde meniaci mierka nemení vlastnosti a pravidlá systému. Príkladom takéhoto javu môže byť fragmentovaný geometrický tvar, rozdelený na časti, z ktorých je každá rovnaká, zmenšená kópia celého geometrického tvaru nazývaný fraktál.

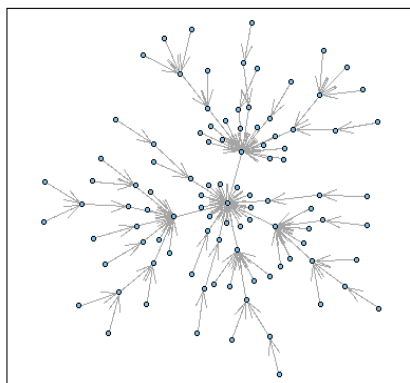
#### 4.4.2 História

Zo štúdií sietí, ako napríklad citačných sietí a vedeckých prác fyzika Dereka de Solla Price z roku 1965 vyplynulo, že reálne siete podliehajú Paretovmu zákonu. Taliansky ekonóm, filozof a politológ Vilfredo Pareto sa zaujímal o distribúciu príjmov v spoločnosti, známy pod menom Paretov princíp alebo aj ako zákon 80/20. Tento zákon hovoril o tom, že 80 % majetku štátu je v rukách 20 % populácie. Na základe tohoto princípu sa snažil dokázať, že rozdelenie príjmov nie je náhodné. Podľa Pareta je hierarchizácia prirodzená a tí, ktorí dosiahnu najvyšší index v postavení v spoločnosti na základe meradla úspechu sa stávajú elitou. Tvrdil, že spoločnosť je pluralistický systém vytvorený z ľudských molekúl, ktoré sú spojené zložitými vzájomnými vzťahmi.

Taktiež si ako záhradkár všimol, že 80 % jeho úrody hrachu pochádza iba z 20 % hrachových luskov. Murphyho zákon managementu hovorí, že 80 % zisku produkuje 20 % zamestnancov, 80 % zločinov pácha 20 % zločincov. Avšak každý systém, ktorý sa riadi paretovým zákonom, má vlastné atribúty a tieto atribúty systémy navzájom odlišujú [4]. Môžeme predpokladať, že reálne siete sa približujú práve modelu bezškálových sietí a riadia sa mocninovým zákonom.

#### 4.4.3 Vlastnosti bezškálových sietí

Bezškálovými sieťami sa začali zaoberať od roku 1999 hlavne fyzik Albert-László Barabási spoločne s kolegami, ktorí skúmali konektivitu internetu. Zistili, že sa niektoré vrcholy vyznačujú rozsiahlym množstvom spojení a niektoré veľmi nízkym. Vrcholy s vysokým počtom stupňov nazvali centrá, obklopené vrcholmi s nižšími stupňami a tieto vrcholy ešte s nižšími. Distribúcia rozdelenia stupňov sa riadi mocninovým zákonom rozobraným v kapitole 2.3.2. Pravdepodobnosť, že vrchol susedí s  $k$  inými vrcholmi je daná vzťahom  $P(k) = k^{(-\alpha)}$ , kde  $\alpha$  je reálny koeficient distribúcie. Exponent  $\alpha$  značí, koľkokrát menej je veľmi obľúbených webových stránok oproti menej obľúbených. Hovorí sa mu aj exponent konektivity. Experimentálne bolo zistené, že u reálnych sietí je  $\alpha$  v rozmedzí  $2 < \alpha < 3$ , nie je to však pravidlom [4]. Na obrázku 6 [19] vidíme vytvorenú bezškálovú sieť.



Obrázok 6: Bezškálová sieť

#### 4.4.4 Rast a preferenčné pripojovanie

Charakteristickými črtami ako rast a preferenčné pripojovanie pre tieto siete definoval Albert-László Barabási ako významný jav absentujúci v náhodných grafoch a ako jav potrebný k vzniku bezškálovej siete. Počet vrcholov v grafe expanduje s časom a to spôsobom, že sa vrchol pripojí už k existujúcim vrcholom, kde je pravdepodobnosť pripojenia daná podelením stupňa  $k_i$  so súčtom všetkých stupňov  $p_i = \frac{k_i}{\sum_j k_j}$ . Preferenčne pripojiť vrchol znamená, že je veľká pravdepodobnosť pripojenia k uzlu s čo najväčším počtom pripojení.

Ďalšou črtou je zdatnosť, ktorá zvyšuje šancu získať nové pripojenia, aby sa eliminovala možnosť, že by mal graf iba jedno centrum, aj keď možné to je. Zdatnosť by sme mohli definovať aj ako miera atraktívnosti uzla pripájať na seba nové uzly v konkurenčnom prostredí. Na zdatné uzly sa pripájajú častejšie nové uzly. Z dvoch uzlov získa väzby rýchlejšie ten, kto má väčšiu zdatnosť. Ak sú však dva uzly rovnako zdatné, starší uzol je vo výhodnejšom postavení pripojiť na seba nové uzly ako mladší [4].

Bezškálové siete sa vyznačujú tvorením zhlukov, kedy vrcholy s nízkym stupňom tvoria podgrafy a tieto podgrafy sú pospájané cez centrá s ešte vyšším stupňom. Ďalšou vlastnosťou bezškálových sietí je vysoká odolnosť voči rozpadu, nakoľko najväčší počet vrcholov má nízky distribučný stupeň a práve preto je pri náhodnom odobraní najvyššia pravdepodobnosť pri odobraní takéhoto uzla najväčšia možná, čo nespôsobí rozpad si-

ete. Pri cielenom útoku na centrá sa rozpadne na množstvo nesúvislých grafov. Preto sú centrá silnou aj slabou oblasťou v sieti [4].

#### 4.4.5 Distribúcia stupňov bezškálových sietí

Pravdepodobnosť, že v sieti bude  $m$  vrcholov so stupňom  $k$  a nebudú existovať žiadne vrcholy s väčším stupňom je daná  $\binom{n}{m} p_k^m (1 - P_k)^{n-m}$ , kde  $P_k$  je kumulatívna pravdepodobnosť distribúcie. Potom pravdepodobnosť  $h_k$ , že získame najvyšší stupeň  $k$  vrchola v sieti je  $h_k = \sum_{m=1}^n \binom{n}{m} p_k^m (1 - P_k)^{n-m} = (p_k + 1 - P_k)^n - (1 - P_k)^n$  a očakávaná hodnota najvyššieho stupňa je  $k_{max} = \sum_k k h_k$  [8].

#### 4.4.6 Priemerná vzdialenosť bezškálových sietí

Priemerná vzdialenosť je menšia ako v náhodných grafoch. Krátke vzdialenosti medzi uzlami su efektívnejšie. Barabásiho model ju definuje ako  $L \sim \frac{\ln(N)}{\ln \ln(N)}$  [18].

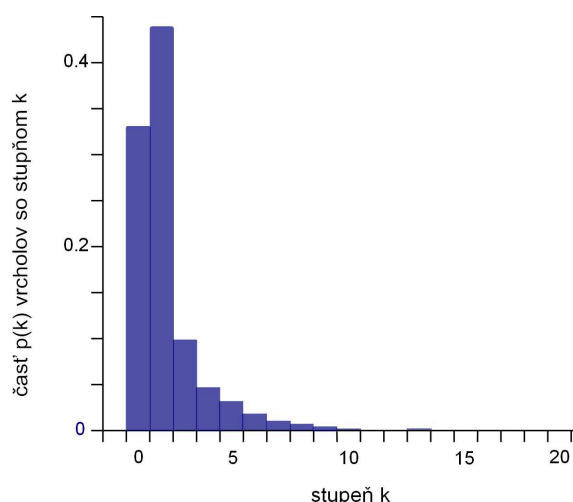
#### 4.4.7 Zhlukovací koeficient bezškálových sietí

Zhlukovací koeficient pre vrchol  $m$  so stupňom  $k$ , je definovaný ako  $C_m = \frac{2n_m}{[k_m(k_m-1)]}$ , kde  $n_m$  je počet hrán vedúce ku  $k_m$  susedom vrchola  $m$  [16]. Zistilo sa, že zhlukovací koeficient je päťkrát vyšší ako u náhodného grafu [18].

## 5 Algoritmy na zistenie modelu siete

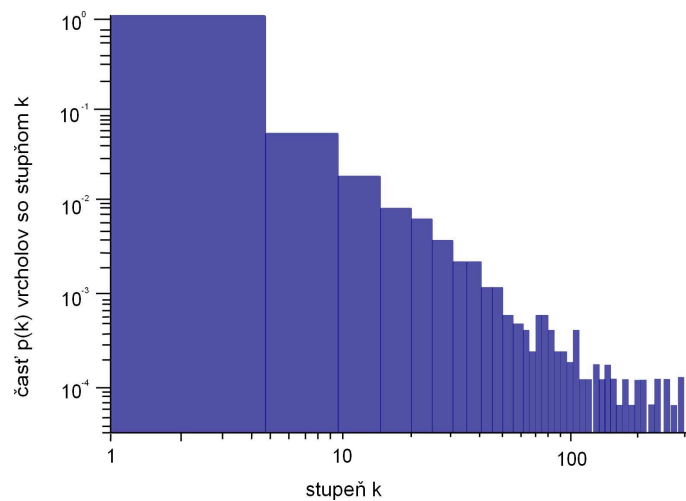
### 5.1 Logaritmický binning

Stratégií, ako vypočítavať vlastnosti bezškálových sietí poznáme viacero. Na obrázku 7 máme možnosť pozorovať stupňovú distribúciu internetu. Vidíme množstvo vrcholov  $n$  s nízkym distribučným stupňom (1, 2, 3) a na konci histogramu nízky počet vrcholov s vysokým distribučným stupňom. Z obrázka 7 vidieť, že značná časť skúmaných dát zostala nezmapovaná, keďže stupeň  $k$  dosahuje hodnotu 20 z celkového počtu 2407. Najväčšie centrum bolo spojené s 12 % ostatnými vrcholmi, čo je približne s každým ôsmym, z celkového počtu 19 956 vrcholov [10].



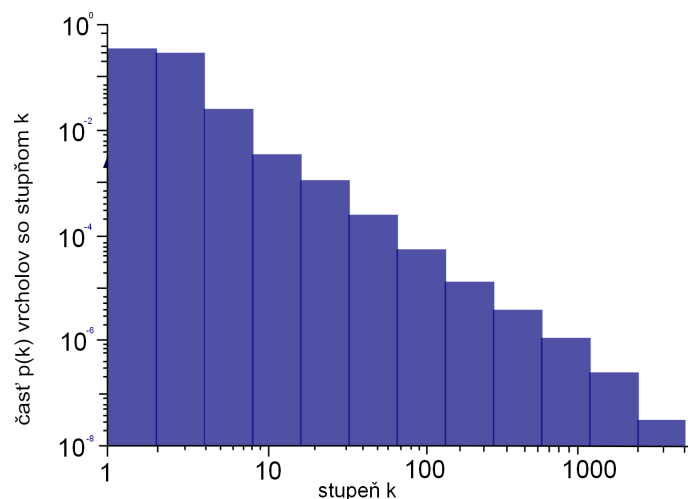
Obrázok 7: Histogram stupňovej distribúcie v lineárnom zobrazení

Jednou z meritórnych metód na určenie bezškálovosti siete je pomocou histogramu v takej logaritmickej mierke, aby do každej oblasti histogramu v určitom intervale pripadol dostatočný počet vzoriek distribučných stupňov, ktorú môžeme vidieť na obrázku 8. Sklon línie je tu viditeľný, avšak koniec na pravej strane, označovaný aj chvost histogramu neumožňuje presne určiť jej sklon. Využijeme preto odlišné veľkosti častí histogramu na lepšie vyšetrenie konca grafu. To zaistí menší šum na konci histogramu a presnejšie vyjadrí sklon línie, ako môžeme vidieť na obrázok 9. Zazrieme tu v logaritmickej mierke nemenné šírky oblastí, zatiaľ čo v lineárnej mierke je každá oblasť širšia o konštantnú zložku  $a$ . Kompromisne definovať odlišné šírky oblastí histogramu je využívané v logaritmickom binningu ako jedným zo spôsobu zistenia modelu siete [10].



Obrázok 8: Histogram stupňovej distribúcie v logaritmickej mierke

Každá šírka oblasti v schéme vždy o niečo viac zahŕňa väčší počet distribučných stupňov  $k$  o konštantnú zložku  $a$  ako predchodca. Vo všeobecnosti sa  $n$ -tá oblasť pokryje v intervale  $a^{(n-1)} \leq k < a^n$  a jej šírka je definovaná  $a^n - a^{(n-1)} = (a - 1)a^{(n-1)}$ . Napr. pre  $a = 2$  spadajú do prvej oblasti všetky vrcholy  $n$  s distribučným stupňom 1 v intervale  $1 \leq k < 2$ . Do druhej oblasti spadajú tie, ktoré spĺnia interval distribučného stupňa  $2 \leq k < 4$ , do tretej oblasti  $4 \leq k < 8$  atď.



Obrázok 9: Histogram stupňovej distribúcie v logaritmickej mierke

Vyjadrenie vrcholov  $n$  so stupňom 0 by nespôsobilo značnú divergenciu histogramu, avšak ich miera sa určuje nezávisle a nezahŕňame ich do schémy v logaritmickej mierke [10].

## 5.2 Kumulatívna funkcia

Kumulatívna funkcia je jedným z ďalších spôsobov ako vizualizovať mocninové rozdelenie.  $P_k$  je pravdepodobnosť, že náhodne vybraný vrchol má distribučný stupeň  $k$  alebo väčší. Pravdepodobnosť je daná  $P_k = \sum_{k'=k}^{\infty} p_{k'}$ . Predpokladajme, že  $P_k$  má mocninové rozdelenie na konci grafu. Povedzme, že  $p_k = Ck^{-\alpha}$  pre  $k \geq k_{min}$  pre nejaké  $k_{min}$ . Potom platí  $P_k = C \sum_{k'=k}^{\infty} k'^{-\alpha} \cong C \int_k^{\infty} k'^{-\alpha} dk' = \frac{C}{\alpha-1} k^{-(\alpha-1)}$  kde sme integrálom približne určili sumu, kde mocninový zákon vykazuje premenlivú funkciu pre rozsiahle  $k$ , pri predpoklade  $\alpha > 1$ , integrál konverguje. Kumulatívna funkcia má exponent  $\alpha$  o 1 menší ako v logaritmickej binningu.

Definujme  $r$  ako počet vrcholov so stupňom vyšším alebo rovnakým ako vybraný  $r$ -tý vrchol. Potom pre pravdepodobnosť, že časť vrcholov s vyšším alebo rovnakým stupňom platí  $P_k = \frac{r}{n}$ . Jednoduchým spôsobom zmapovania kumulatívnej funkcie je zoradiť stupne vrcholov od najvyššieho po najmenší a každému priradzovať hodnoty  $r$  v rozsahu od 1 až po  $n$ . Vykreslenia jednotlivých  $\frac{r_i}{n}$  ako funkciu stupňa  $k_i$  s vrcholmi v poradí  $r$  získame vykreslenie kumulatívnej funkcie [10].

## 5.3 Maximálny vierohodný odhad

Použitím vzorca na zmapovanie sklonu línie je jedným z ďalších spôsobov, ktorý môžeme využiť. Exponent  $\alpha$  je daný  $\alpha = 1 + n[\sum_i \ln \frac{k_i}{k_{min}-\frac{1}{2}}]^{-1}$ , kde  $k_{min}$  je zvolený minimálny stupeň,  $n$  je počet vrcholov väčších alebo rovných  $k_{min}$ . Celková suma platí len pre vrcholy  $k \geq k_{min}$ . Štatistickú chybu, ktorá vzniká použitím predchádzajúcej formuly vyjadríme  $\sigma = \sqrt{n[\sum_i \ln \frac{k_i}{k_{min}-\frac{1}{2}}]^{-1}} = \frac{\alpha-1}{\sqrt{n}}$  [10].



## 6 Experimentálna časť

Hlavným cieľom experimentálnej časti tejto práce je uskutočniť pokusy za pomoci vlastnej implementácie programu na rôznych dátových kolekciách, ktoré predstavujú reálne siete. Na základe distribúcie stupňov vrcholov určujeme modely, viažuce sa na konkrétne pokusné reálne siete. Výsledné dáta sú prezentované grafmi, vyhotovených v programe Gnuplot vo verzii 4.6 [15], spoločne so stručným popisom. Počítačová aplikácia, ktorá je súčasťou práce je schopná zistiť mocninný exponent  $\alpha$  prostredníctvom troch metód:

1. maximálneho vierohodného odhadu,
2. logaritmického binningu,
3. kumulatívnej funkcie.

Výsledné dáta prezentované grafmi, boli vypočítané touto počítačovou aplikáciou zo vstupných dát v požadovanom formáte očistených od duplicít. Mocninný exponent  $\alpha$  zistený tromi rôznymi metódami nám umožňuje porovnať presnosť týchto metód. Maximálny vierohodný odhad sa považuje za najpresnejší spôsob výpočtu mocninného exponentu  $\alpha$ . Za ním nasleduje spôsob výpočtu kumulatívnou funkciou a za najmenej presný sa pokladá logaritmický binning.

### 6.1 Analýza požiadavkov

Aplikácia prijíma na vstup textový súbor obsahujúci vrcholy a ich hrany. V takomto vstupnom súbore je vrchol realizovaný ako číslo a hranu predstavujú dva vrcholy v jednom riadku oddelené oddeľovačom a to medzerami, tabulátormi alebo iným oddeľovačom. V aplikácii sa vyberá možnosť výpočtov pre orientovaný, či neorientovaný graf. Výstupom sú hodnoty exponentu  $\alpha$ , zaokrúhlené na tri desatinné miesta.

#### Funkčné požiadavky

1. Aplikácia načíta vstupný súbor.
2. Aplikácia ukladá do textových súborov stupňovú distribúciu pre každý vrchol a pre jednotlivé stupne počty vrcholov s týmto stupňom.
3. Aplikácia vyhodí duplicity pre orientovaný a neorientovaný graf.

4. Aplikácia počíta exponent  $\alpha$  troma metódami a zobrazí výsledky s presnosťou na tri desatinné miesta.

### Ostatné požiadavky

1. Aplikácia podporuje textové súbory v požadovanom formáte dát.

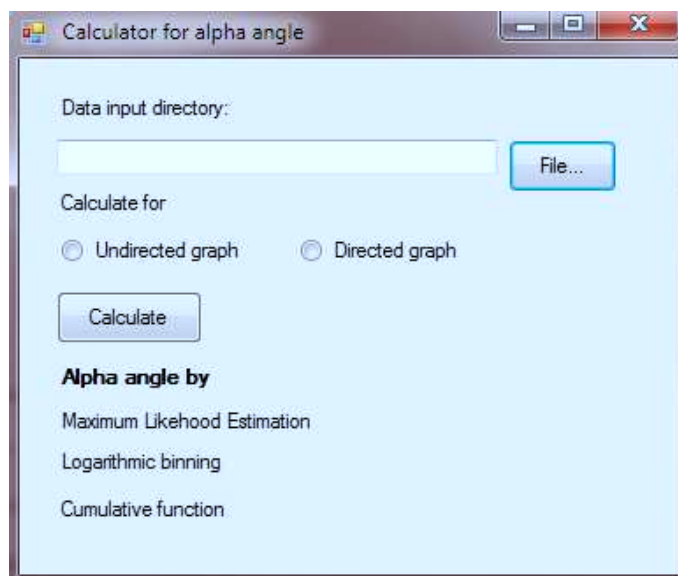
## 6.2 Návrh

V prílohe A je zobrazený triedny diagram aplikácie. Hlavnou triedou, ktorá sprostredkuje komunikáciu medzi ostatnými triedami je trieda `Form1`, ktorá komunikuje s užívateľom. Užívateľ vo formulári zadá cestu k textovému súboru, pre ktorý majú byť zrealizované výpočty exponentu  $\alpha$  a zvolí typ grafu pomocou riadiaceho prvku `RadioButton`, pomocou ktorého priraduje atribútu `what_checked` typ grafu.

Triedy `UndirectedGraph` a `DirectedGraph` obsahujú základné operácie pre prácu so súbormi, výpočtové operácie stupňovej distribúcie vrcholov, využívajúce generické kolekcie typov `Dictionary<int, List<int>>`, `HashSet<string>`, `List<string>`. Sú vybrané také typy kolekcí, ktoré sú prijateľné, nie najrýchlejšie vzhľadom na časové nároky. Triedy `LogBin`, `MaximumLikeHoodEstimation`, `CumulativeFunction` obsahujú vlastné operácie na výpočet exponentu  $\alpha$ . Statickú triedu `Calculations` využíva trieda `LogBin` na kalkuláciu priemerov centroidov, ktorá by mohla byť v budúcnosti rozšírená v prípade, že by bolo potreba kalkulovať exponent  $\alpha$  ďalšími spôsobmi, kde by sa metóda `DoAverage` v tejto triede mohla ešte využiť.

### 6.2.1 Užívateľské rozhranie

Užívateľské rozhranie je tvorené jednoduchým oknom z knižnice *WinForms*, obsahujúci riadiaci prvok `textBox`, ktorý zobrazuje cestu textového súboru. Po stlačení tlačidla "File...", je vytvorený objekt typu `OpenFileDialog`, ktorý zabezpečuje otvorenie súboru. Nechýba možnosť výberu výpočtu pre orientovaný a neorientovaný graf. Po stlačení tlačidla *Calculate* začne prebiehať výpočet exponentov  $\alpha$  všetkými troma metódami postupne. Aplikácia upozorňuje, že výpočet prebieha, pomocou zmeneného kurzoru myši na "wait". Po ukončení behu výpočtov exponentov sa zobrazia vypočítané hodnoty spoločne s hláškou "Done." vedľa tlačidla *Calculate*. Formulár užívateľského rozhrania vidíme na obrázku 10.



Obrázok 10: Užívateľské rozhranie

### 6.3 Implementácia

Triedy `UndirectedGraph` a `DirectedGraph` obsahujú operácie pre čítanie z textového súboru, následné pridávanie do zoznamu s vyhadzovaním duplícít, a to rozdielnym spôsobom pre každú triedu zvlášť. Prvotnou myšlienkou a zotrvajúcou i keď nie najlepšou, bolo dáta ukladať do kolekcie v podobe reťazca, preto je v triede `UndirectedGraph` implementovaná metóda `ReverseString`, ktorá je účelná pre vyhadzovanie duplícít typu 1 - 2, 2 - 1 a nevhodná pre triedu `DirectedGraph`.

---

```
public string ReverseString(string s)
{
    string [] word = SplitWords(s);
    string str = word[1] + ";" + word[0];
    return str;
}
```

---

Výpis 1: Metóda `ReverseString` vracajúca opak reťazca

Metóda `SplitWords` slúži na prístup k jednotlivým hodnotám v riadkoch využívajúc regulárneho výrazu.

---

```

public string [] SplitWords(string word)
{
    return Regex.Split(word, @"\W+");
}

```

---

Výpis 2: Metóda SplitWords vracajúca hodnotu v riadku

Vzhľadom na vysoké časové nároky na počítanie distribúcie stupňov vo forme zoznamu reťazcov, sa zhotovila nová implementácia počítania stupňovej distribúcie a preto bol pre tento účel zvolený objekt typu `Dictionary` s kľúčom typu `int`, a hodnotami typu generického číselného zoznamu.

```

connexion = new Dictionary<int, List<int>>();

```

Metóda `AddToConnexion()` do tohto objektu typu `Dictionary` pridá vrcholy a ich spojenia s inými vrcholmi. Pomocou ďalšej metódy `DegreeOfEdges` sa vypočíta stupňová distribúcia a uloží sa do nového zoznamu reťazcov `verticesDistribution`, ktorý sa v metóde `WriterVerticeDistribution` zapisuje do textového súboru.

---

```

public void DegreeOfEdges()
{
    Dictionary<int, List<int>>.KeyCollection keycollection = connexion.Keys;
    int i = 0;
    foreach (int key in keycollection)
    {
        stupen[0][i] = key;
        stupen[1][i] = connexion[key].Count;
        i++;
        verticesDistribution .Add(key + " - " +
            connexion[key].Count);
    }
}

```

---

Výpis 3: Metóda DegreeOfEdges na výpočet stupňovej distribúcie

Z kolekcie `verticesDistribution`, v ktorom sú dáta deklarované ako vrchol - počet stupňov tohto vrchola, sa v metóde `ComputeDegreeQuantity` naplní kolekcia typu `List<string> degreesQuantity`, kde sú dáta uložené ako stupeň - počet vrcholov

s týmto stupňom. V metóde `WriterDegreesQuantity` sa kolekcia `degreesQuantity` zapisuje do textového súboru.

---

```
public void WriterDegreesQuantity()
{
    using (StreamWriter writer = new StreamWriter ("degreesQuantityUndirectedGraph.txt", false))
    {
        foreach (string s in degreesQuantity)
        {
            writer.WriteLine(s);
        }
    }
}
```

---

#### Výpis 4: Metóda `WriterDegreesQuantity` zapisuje do súboru

Pre volanie týchto základných operácií uvedených vyššie bola zhotovená metóda nazvaná `GetdegreesQuantity()` vracajúca kolekciu `degreesQuantity`, ktorú využíva objekt vytvorený v triede `Form1`. Metódy `GetLogBin()`, `GetCumulativeFunction()`, `GetMaximumLikelihoodEstimation()` vytvárajú objekty, ktoré získajú hodnotu  $\alpha$  pre kolekciu `degreesQuantity`.

Trieda `MaximumLikelihoodEstimation` obsahuje metódu `GetAlpha()` s jedným parametrom a vracajúcou hodnotou typu *double*, v ktorej je riešený exponent  $\alpha$ . Pre výpočet exponentu  $\alpha$  metódou logaritmickeho binningu v triede `LogBin` je potrebné vyčíslieť dva centroidy, pre ktoré sa naimplementovali metódy `ListForFirstCentroid()`, `ListForSecondCentroid()` používajúce kolekcie, do ktorých sa ukladajú hodnoty z intervalov potrebné pre výpočet súradníc centroidov. Kolekcie sú zlogaritmované a následne sa pomocou triedy `Calculations`, obsahujúca metódu `DoAverage`, počítajú priemery hodnôt, predstavujúce súradnice centroidov. Metóda `CalculateDirection()` vyhodnocuje smernicu určenú týmito centroidmi, ktorá je zároveň exponentom  $\alpha$ .

---

```
public void CalculateDirection()
{
    smernica_average = Math.Abs(((y2_average) - (y_average)) / ((x2_average) - (x_average)));
    posun_average = (y_average) - (smernica_average * x_average);
}
```

---

---

#### Výpis 5: Metóda CalculateDirection na výpočet exponentu

Trieda `CumulativeFunction` využíva princíp výpočtu v triede `LogBin`, avšak kolekcia dát, z ktorého sa vyhodnocuje exponent je odlišná, ktorú si trieda sama počíta v metóde `Cumulativefunction()`.

#### Použité technológie

1. programovací jazyk `c#`,
2. .NET Framework 4,
3. Visual Studio 2010 Ultimate.

#### Spustenie aplikácie

Pre spustenie aplikácie je potrebné mať nainštalovaný .NET Framework 4. Je nutné vygenerovať binárny súbor `Calculator.exe`, ktorý aplikáciu spustí. K tomu je nutné mať nainštalovaný prekladač, ktorý si pomocou súboru `Calculator.sln` zostaví a preloží aplikáciu a vytvorí spustiteľný súbor.

## 6.4 Experimenty nad vybranými datovými kolekciami

Dôležitou súčasťou práce je prevedenie experimentov nad vybranými dátovými kolekciami. Nakoľko sú siete dynamickými systémami, neustále sa meniace v čase a kolekcie dát, ktoré máme k dispozícii sú staršieho dátumu, tak nezodpovedajú súčasnému stavu reálnych sietí. V nasledujúcej tabuľke máme možnosť vidieť exponenty  $\alpha$  z jednotlivých kolekcí. Pre orientované grafy je zápis exponentov v tabuľke : vstupné/výstupné hrany. Hodnota  $x_{min}$  pre MLE všetkých skúmaných kolekcí bola zvolená 1.

Kolekcia	MLE	Kumulatívna funkcia	Logaritmický binning
Wikipedia češtiny	1,257/1,251	1,479/1,398	1,470/1,506
Wikipedia taliančiny	1,232/1,216	1,434/1,144	1,547/1,529
Wikipedia francúzštiny	1,205/1,183	1,315/1,014	1,705/1,667
Spolupráca autorov	1,374	2,963	3,262
Autonómne systémy	1,241	1,145	1,513

Tabuľka 1: Tabuľka exponentov  $\alpha$  z kolekcí dát

### 6.4.1 Kolekcia dát Wikipedie v češtine

Obrázok 11 ukazuje v lineárnom zobrazení množstvo odkazov na stránkach, na ktoré sa odkazuje v českej encyklopédii Wikipedia. Najväčší počet dosiahnutých odkazov na stránke, na ktorú sa odkazuje bolo 674, kým 522 stránok je počet, na ktoré sa odkazovala práve jedna stránka. Celkový počet všetkých stránok, na ktoré sa odkazuje bolo 4462 (počet vrcholov). Počet všetkých odkazov medzi stránkami sa vyčíslil na 51 687 (počet hrán). Na obrázku 12 prezentujeme tie isté dáta v logaritmickom zobrazení. Dáta z kumulatívnej funkcie vstupných dát môžeme vidieť na obrázku 13.

Celkový počet všetkých stránok, ktoré odkazujú na iné stránky bol 4 919. Maximálny počet odkazov, ktoré stránka odkazovala na iné stránky bol 735. Jeden odkaz, ktorý odkazoval na inú stránku malo 840 stránok. Obrázok 14 ukazuje v lineárnom zobrazení množstvo odkazov stránok, ktoré odkazujú v českej encyklopédii Wikipedia. Obrázok 15 prezentuje tie isté dáta v logaritmickom zobrazení. Dáta z kumulatívnej funkcie výstupných dát môžeme vidieť na obrázku 16.

#### 6.4.2 Kolekcia dát Wikipedie v taliančine

Obrázok 17 ukazuje v lineárnom zobrazení množstvo odkazov na stránkach, na ktoré sa odkazuje v talianskej encyklopédii Wikipedia. Najväčší počet dosiahnutých odkazov na stránke, na ktorú sa odkazuje bolo 1 252, kým 1 026 stránok je počet, na ktoré sa odkazovala práve jedna stránka. Celkový počet všetkých stránok, na ktoré sa odkazuje bolo 11 628 (počet vrcholov). Počet všetkých odkazov medzi stránkami sa vyčíslil na 164 640 (počet hrán). Na obrázku 18 prezentujeme tie isté dáta v logaritmickej mierke. Dáta z kumulatívnej funkcie vstupných dát môžeme vidieť na obrázku 19.

Celkový počet všetkých stránok, ktoré odkazujú bol 12 430. Maximálny počet odkazov, ktoré stránka odkazovala na iné stránky bol 1 746. Jeden odkaz, ktorý odkazoval na inú stránku malo 2 949 stránok. Obrázok 20 ukazuje v lineárnom zobrazení množstvo odkazov stránok, ktoré odkazujú v českej encyklopédii Wikipedia. Obrázok 21 prezentuje tie isté dáta v logaritmickej mierke. Dáta z kumulatívnej funkcie výstupných dát môžeme vidieť na obrázku 22.

#### 6.4.3 Kolekcia dát Wikipedie vo francúzštine

Obrázok 23 ukazuje v lineárnom zobrazení množstvo odkazov na stránkach, na ktoré sa odkazuje vo francúzskej encyklopédii Wikipedia. Najväčší počet dosiahnutých odkazov na stránke, na ktorú sa odkazuje (čiže stránok, koľko sa na túto stránku odkazovali) bolo 2 250, kým 3 865 stránok je počet, na ktoré sa odkazovala práve jedna stránka. Celkový počet všetkých stránok, na ktoré sa odkazuje bolo 39 461 (počet vrcholov). Počet všetkých odkazov medzi stránkami sa vyčíslil na 702 245 (počet hrán). Na obrázku 24 prezentujeme tie isté dáta v logaritmickej mierke. Dáta z kumulatívnej funkcie vstupných dát môžeme vidieť na obrázku 25.

Celkový počet všetkých stránok, ktoré odkazujú bol 45 502. Maximálny počet odkazov, ktoré stránka odkazovala na iné stránky bol 7 570. Jeden odkaz, ktorý odkazoval na inú stránku malo 10 228 stránok. Obrázok 26 ukazuje v lineárnom zobrazení množstvo odkazov stránok, ktoré odkazujú v českej encyklopédii Wikipedia. Obrázok 27 prezentuje tie isté dáta v logaritmickej mierke. Dáta z kumulatívnej funkcie výstupných dát môžeme vidieť na obrázku 28.

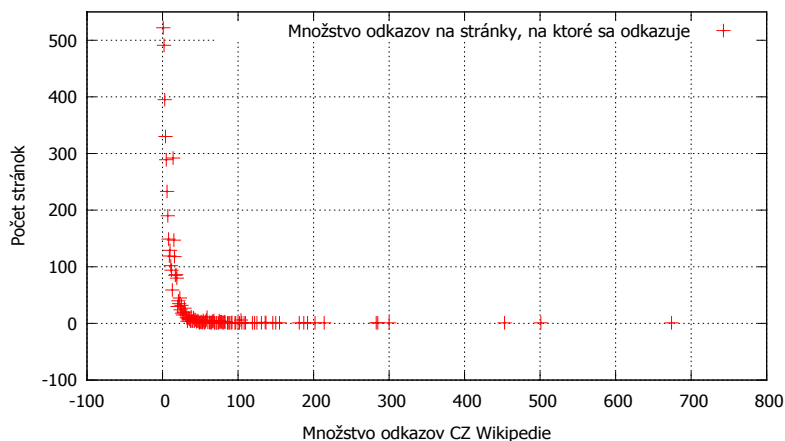


#### 6.4.4 Kolekcia dát reprezentujúcich spoločne publikujúcich autorov

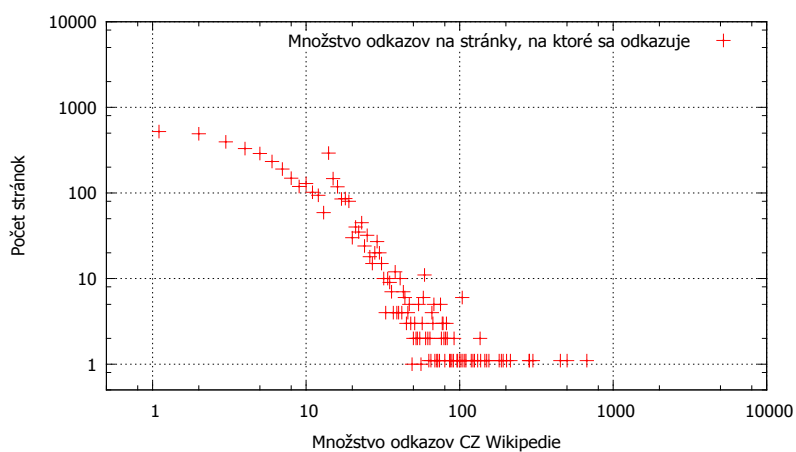
Obrázok 29 znázorňuje v lineárnom zobrazení množstvo autorov spolupracujúcich na autorskom diele v oblasti IT. Najväčší počet spolupracovníkov na diele dosiahol počet 51 ľudí, kým 24 400 diel malo 1 spolupracovníka z celkového počtu 55 875 spolupracovníkov (počet vrcholov). Počet spojení medzi stránkami sa vyčíslil na 135 708 (počet hrán). Na obrázku 30 vidíme v logaritmickej mierke dáta predchádzajúceho grafu. Dáta získane kumulatívnou funkciou máme možnosť vidieť na obrázku 31. Exponent dosiahol hodnotu 2,963.

#### 6.4.5 Kolekcia dát autonómnych systémov

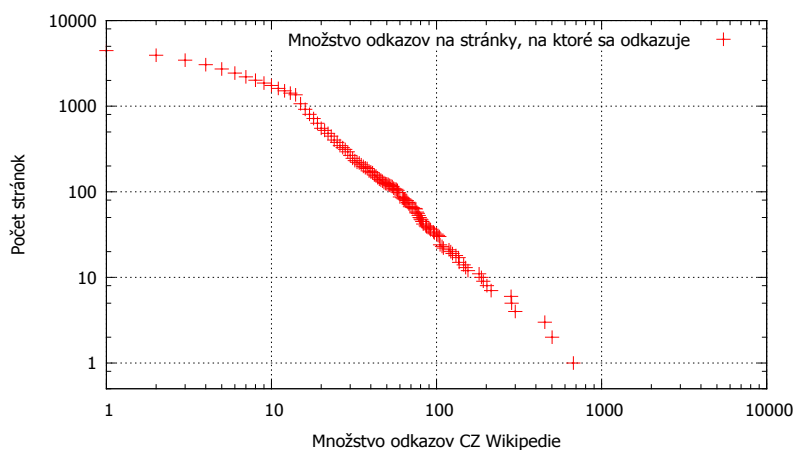
Dáta z poslednej kolekcie znázorňuje obrázok 32 v lineárnom zobrazení a reprezentujú autonómne systémy Internetu. Sú to skupiny počítačov, pod spoločnou správou, ktoré reprezentujú voči Internetu spoločnú routovaciu politiku. S počtom 2 566 spojení s inými autonómnyimi systémami dosiahol autonómny systém maximum. V spojení s dvoma inými autonómnyimi systémami dosiahlo až 5 471 autonómnych systémov z celkového počtu 12 694 (počet vrcholov) autonómnych systémov. Hodnota 53 118 udáva počet všetkých spojení medzi autonómnyimi systémami (počet hrán). Dáta z predchádzajúceho grafu sú na obrázku 33 vykreslené v logaritmickej mierke. Dáta získané kumulatívnou funkciou máme možnosť vidieť na obrázku 34.



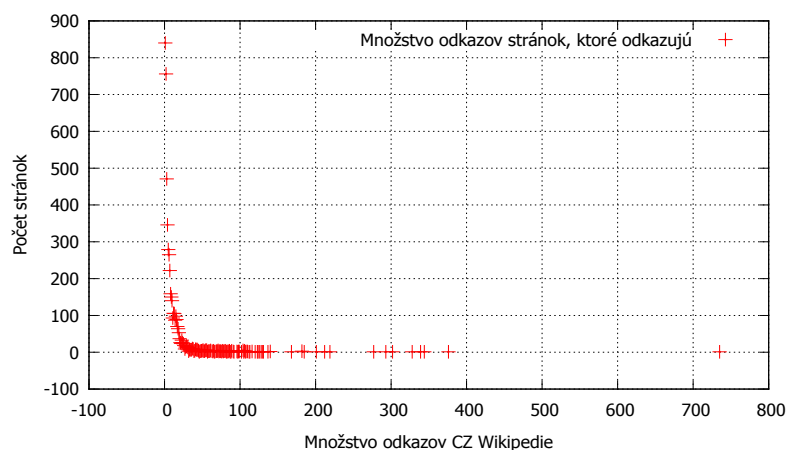
Obrázok 11: Wikipedia češtiny v lineárnom zobrazení



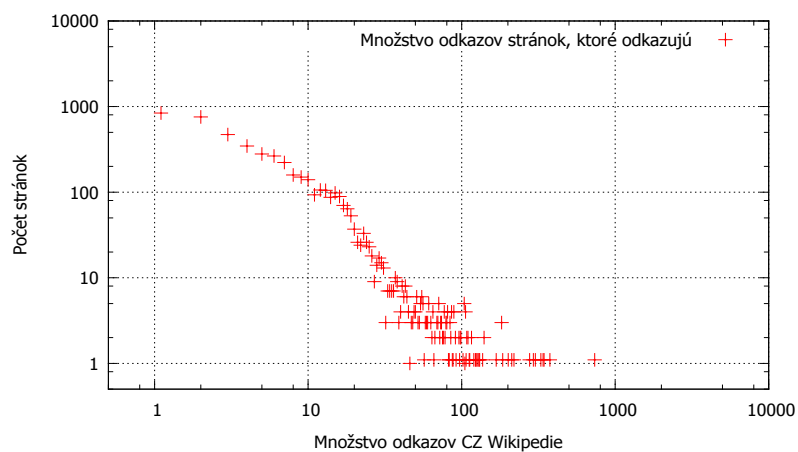
Obrázok 12: Wikipedia češtiny v logaritmickom zobrazení



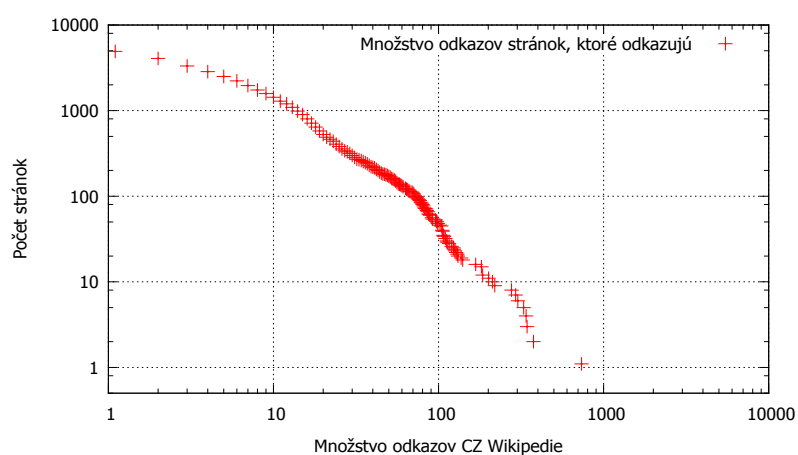
Obrázok 13: Wikipedia češtiny v kumulatívnej funkcii



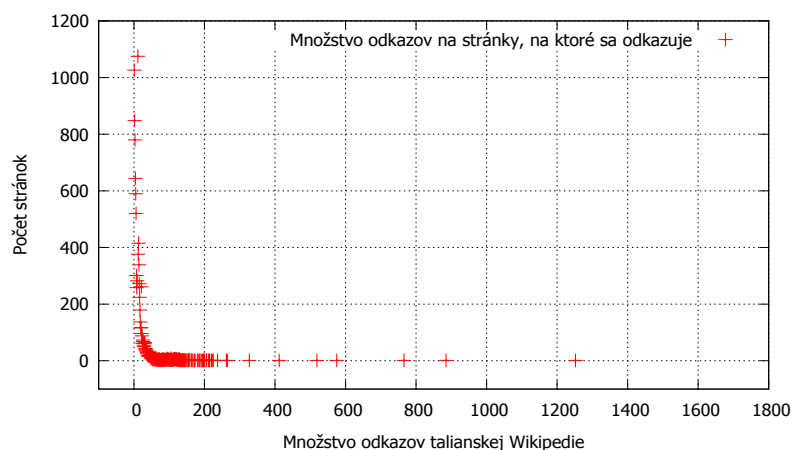
Obrázok 14: Wikipedia češtiny v lineárnom zobrazení



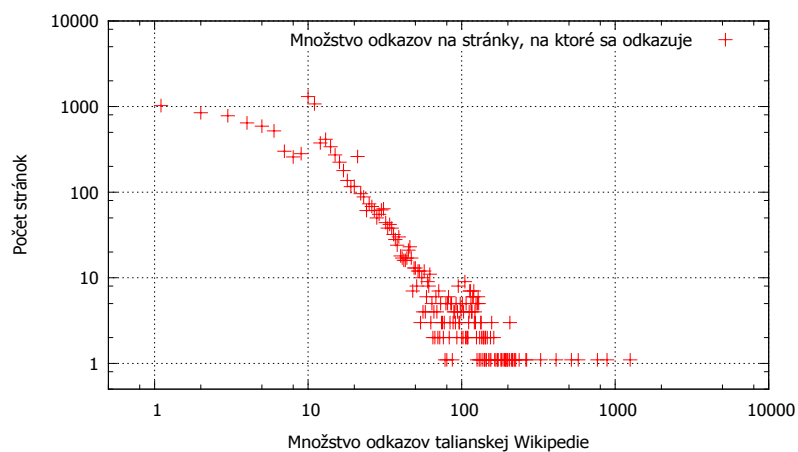
Obrázok 15: Wikipedia češtiny v logaritmickom zobrazení



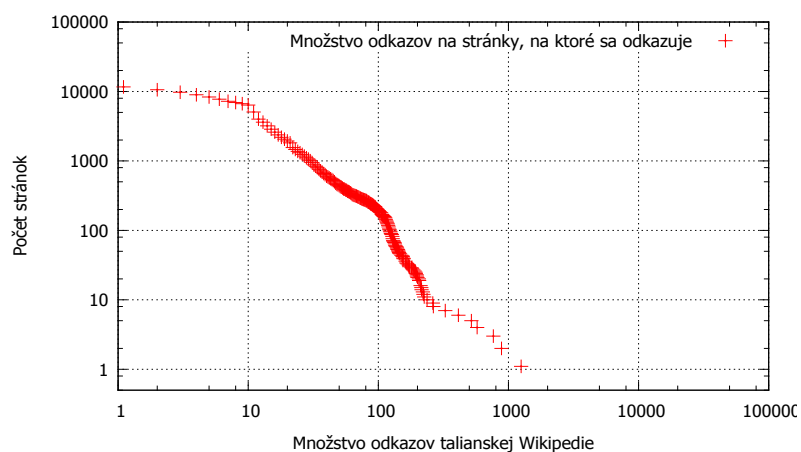
Obrázok 16: Wikipedia češtiny v kumulatívnej funkcii



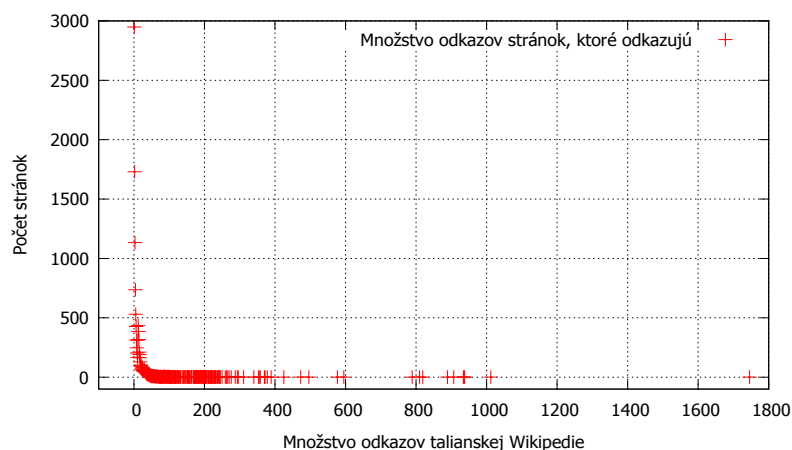
Obrázok 17: Wikipedia taliančiny v lineárnom zobrazení



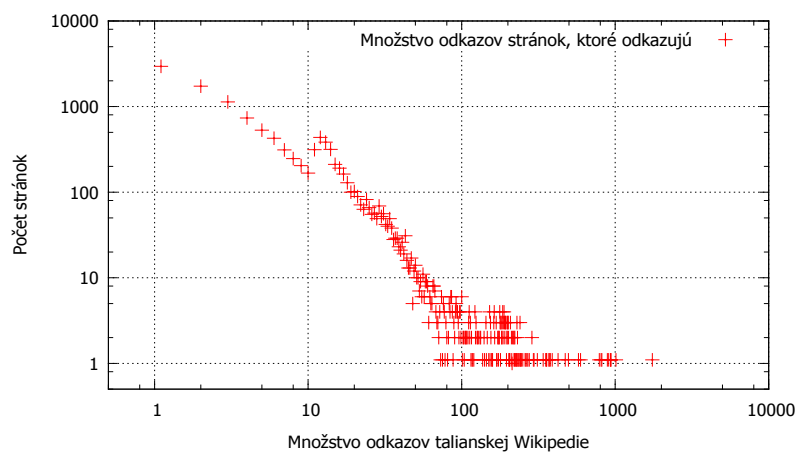
Obrázok 18: Wikipedia taliančiny v logaritmickom zobrazení



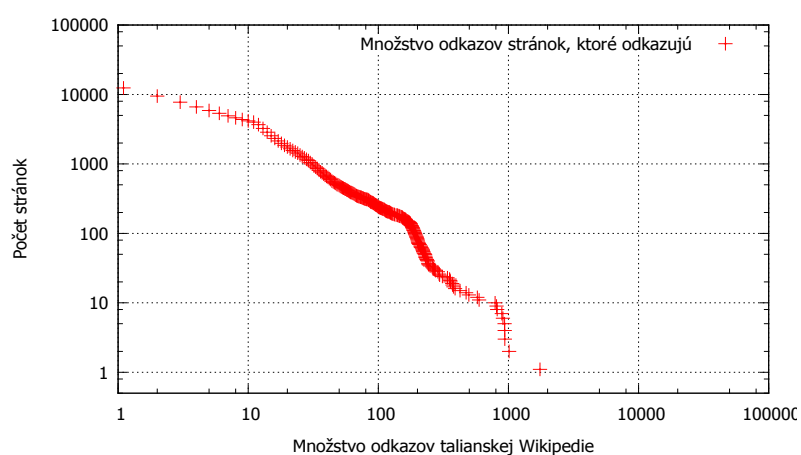
Obrázok 19: Wikipedia taliančiny v kumulatívnej funkcii



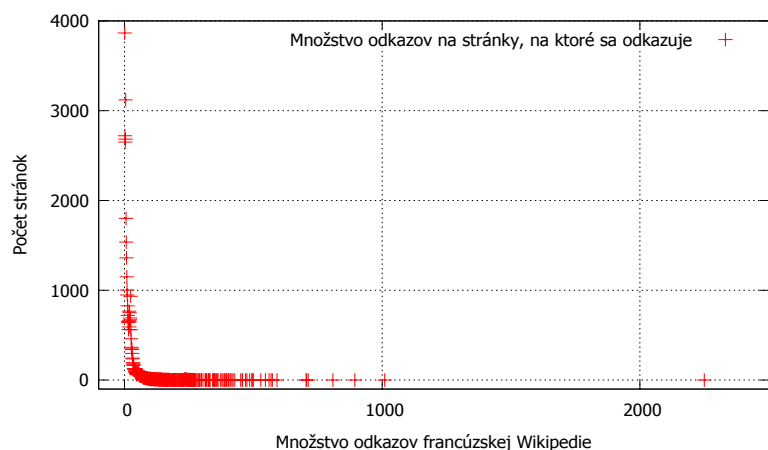
Obrázok 20: Wikipedia taliančiny v lineárnom zobrazení



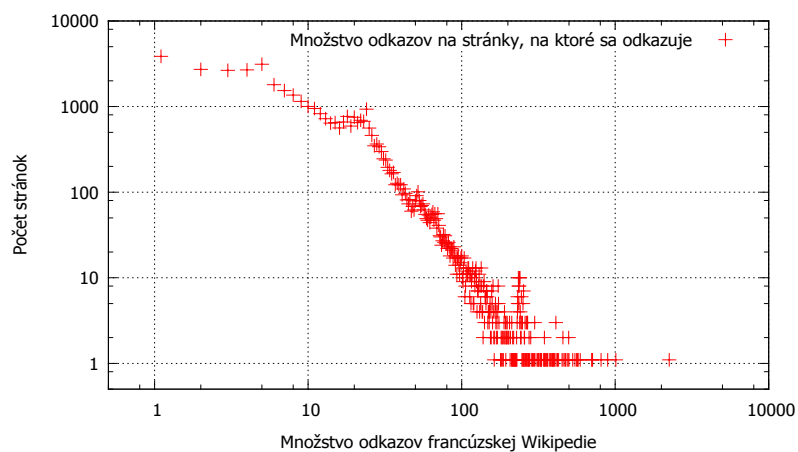
Obrázok 21: Wikipedia taliančiny v logaritmickom zobrazení



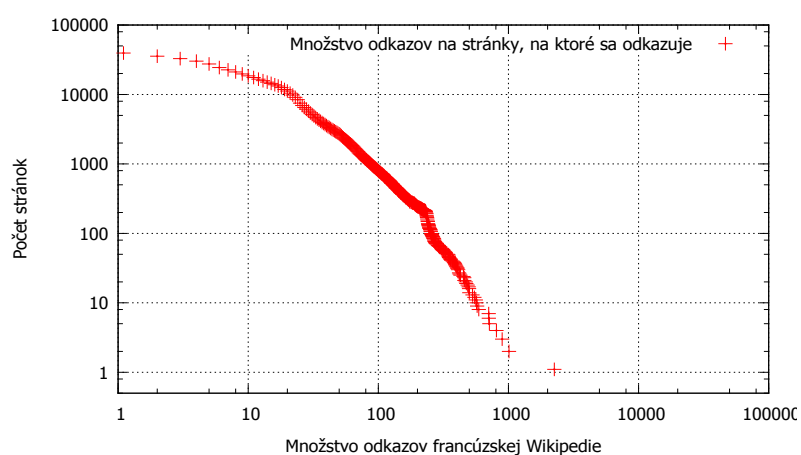
Obrázok 22: Wikipedia taliančiny v kumulatívnej funkcii



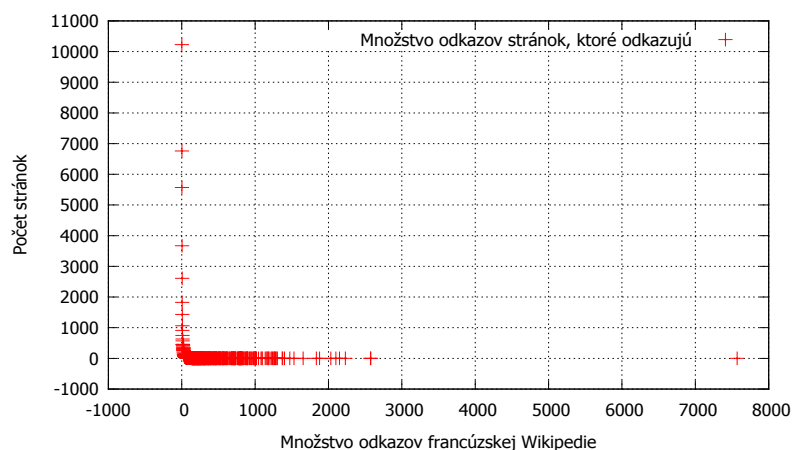
Obrázok 23: Wikipedia francúzštiny v lineárnom zobrazení



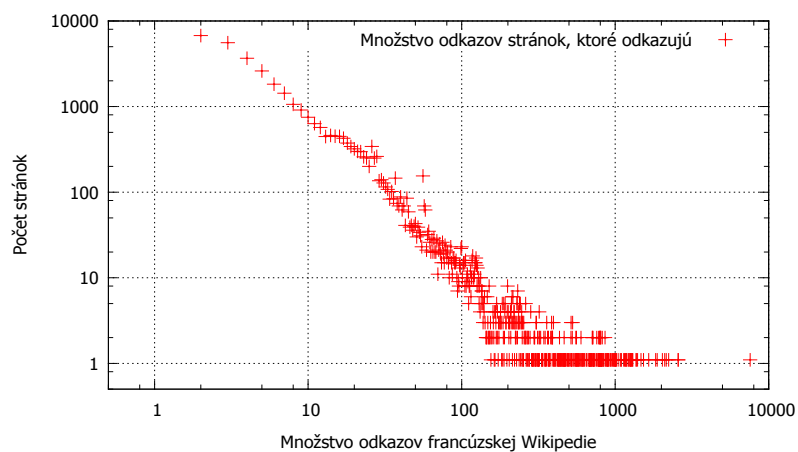
Obrázok 24: Wikipedia francúzštiny v logaritmickom zobrazení



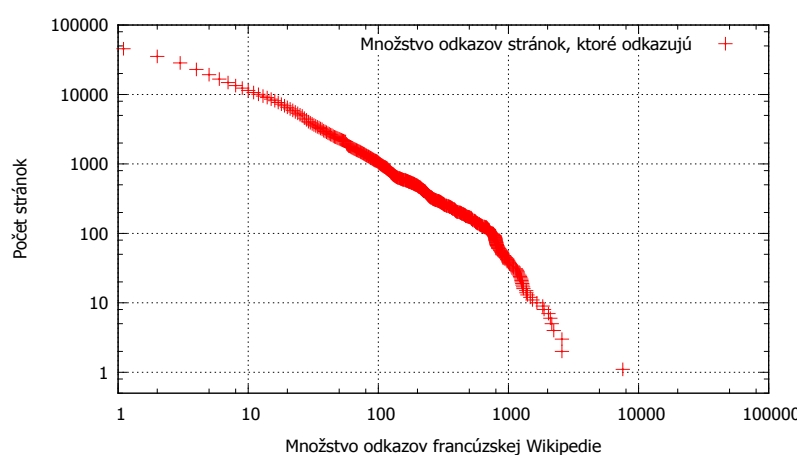
Obrázok 25: Wikipedia francúzštiny v kumulatívnej funkcii



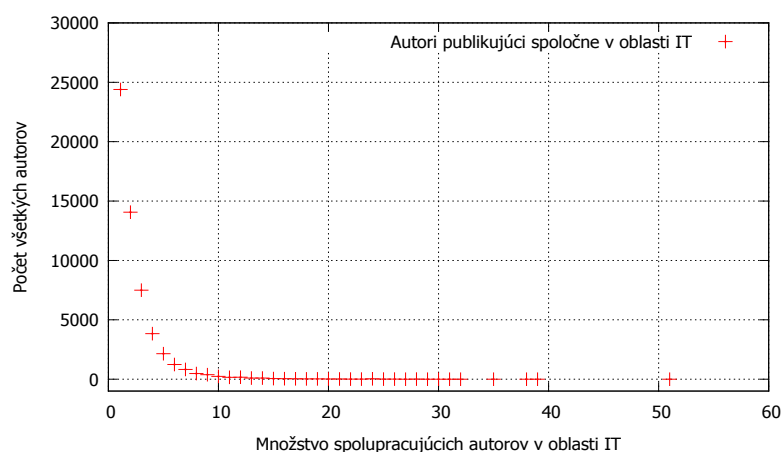
Obrázok 26: Wikipedia francúzštiny v lineárnom zobrazení



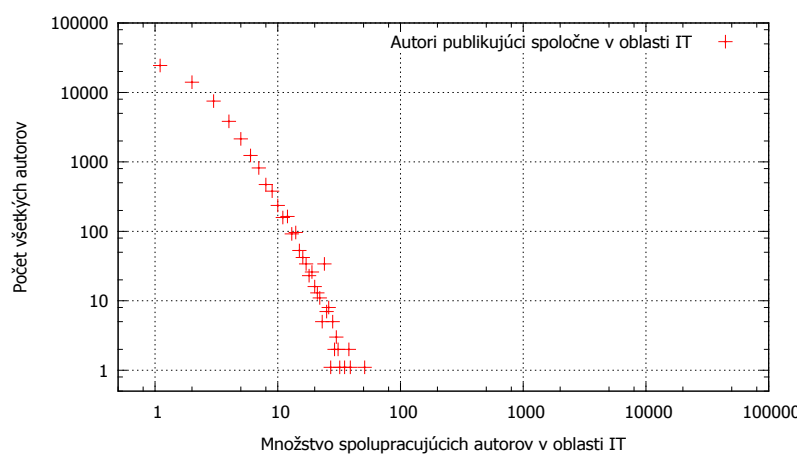
Obrázok 27: Wikipedia francúzštiny v logaritmickom zobrazení



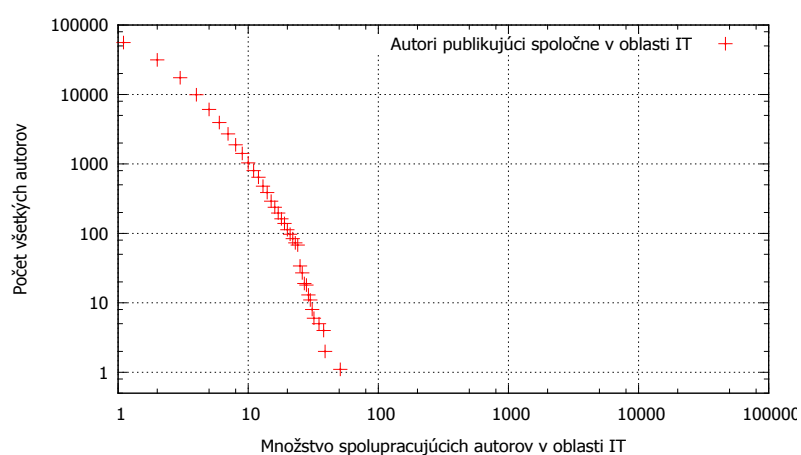
Obrázok 28: Wikipedia francúzštiny v kumulatívnej funkcii



Obrázok 29: Spolupracujúci autori v lineárnom zobrazení

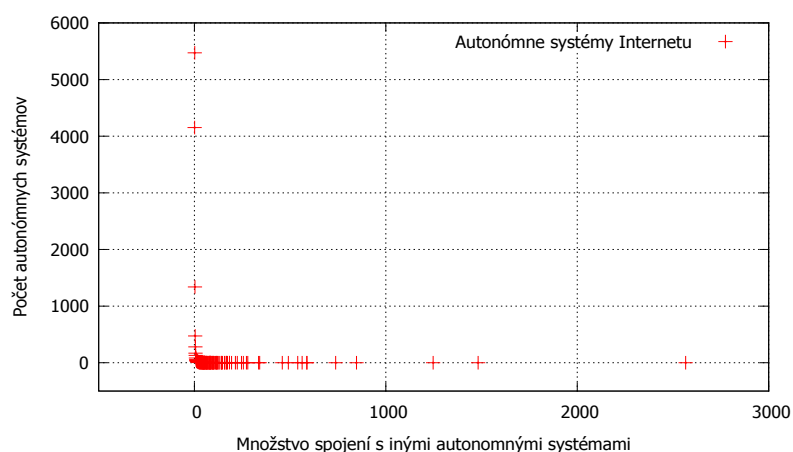


Obrázok 30: Spolupracujúci autori v logaritmickom zobrazení

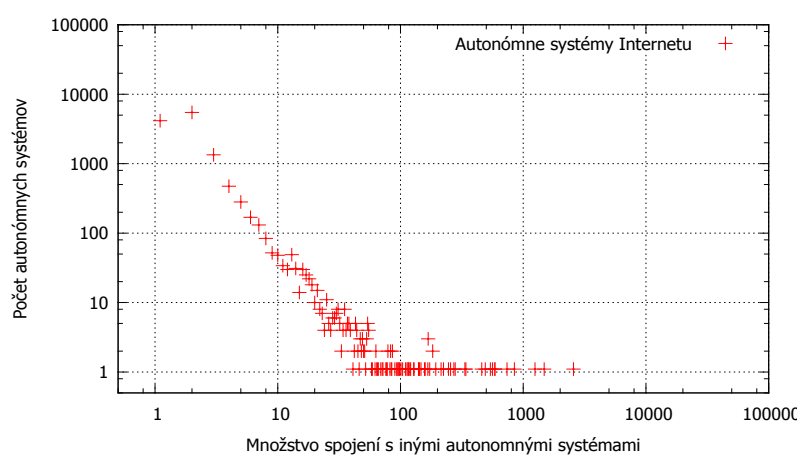


Obrázok 31: Spolupracujúci autori v kumulatívnej funkcii

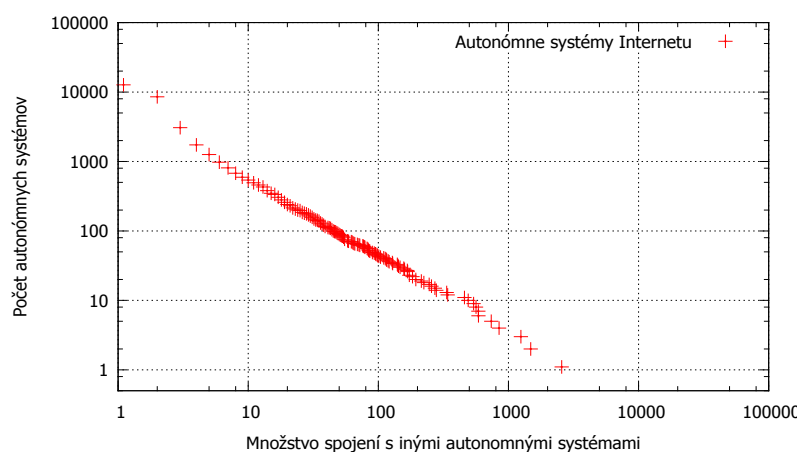




Obrázok 32: Autonómne systémy v lineárnom zobrazení



Obrázok 33: Autonómne systémy v logaritmickej zobrazení



Obrázok 34: Autonómne systémy v kumulatívnej funkcii

## 7 Závěr

Popísať a následne za pomoci vlastnej aplikácie a vybraných dátových kolekcií rozlíšiť základné modely sietí pomocou charakteristických vlastností, bolo hlavnou témou tejto práce. Naimplementovaná aplikácia pomáha skúmať štruktúru reálnych sietí hlavne na základe distribúcie stupňov. Vzájomne odlišné siete, ako siete webových odkazov, či autonómne systémy sa vo svojej štruktúre podobajú, grafy v práci sú jasným príkladom a prichádzame k rovnakému záveru. V týchto sieťach sa tvoria centrá s vysokým počtom napojení a majú tendenciu riadiť sa mocninovým zákonom. Všetky tieto skúmané siete dosiahli exponent  $\alpha$  pomocou najpresnejšej metódy MLE menší než 3, čo je pre bezškálové siete typické. Na základe vlastnosti distribúcie stupňov sme zistili, že skúmané siete patria medzi bezškálové siete. Hlavný cieľ tejto práce sa podarilo naplniť.

Napriek rôznorodosti a s odstupom času môžeme pozorovať nárast počtu uzlov v týchto kolekciách. Je zaznamenaný nárast autonómnych systémov, kde počet registrácií 16-bit ASN dávno presiahlo možný počet, kým naša pokusná dátová kolekcia obsahovala 12 694 autonómnych systémov. ASN je jedinečné číslo pridelené novému autonómnemu systému. Rovnaké pravidlá rastu (s inou intenzitou) riadia i sieť odkazov vo Wikipediách, či sieť spoločne publikujúcich autorov v oblasti IT. Takéto zložité siete sú neustále vo vývoji, nie sú statické.

Vytvorený program, ktorý meria exponent  $\alpha$  tromi spôsobmi a ktorý zaznamenáva distribúciu stupňov sa v budúcnosti dá optimalizovať tak, aby výpočet distribúcie bol rýchlejší. Doimplementovať sa dá množstvo nových rozšírení, pre výpočet exponentu  $\alpha$  prichádza do úvahy možnosť riešiť ho pomocou metódy najmenších štvorcov. Rozšíriť aplikáciu o vykresľovanie grafov, o zobrazovanie najvyšších a najnižších distribučných stupňov a rôzne ďalšie zaujímavé informácie, ktoré v tomto programe treba prácne extrahovať z textového súboru, sú taktiež možné.

## 8 Literatura

- [1] *Komplexné siete okolo nás a v nás*  
Dostupné z: [http://www.gjh.sk/files/prednasky/prezent/16\\_nather.pdf](http://www.gjh.sk/files/prednasky/prezent/16_nather.pdf) [cit. 2013-04-25].
- [2] *Funkčné topológie mozgu*  
Dostupné z: <http://dai.fmph.uniba.sk/events/kuz2009/prispevky-pdf/nather.pdf> [cit. 2013-04-25].
- [3] *Jon Kleinberg*  
Dostupné z: <http://www.cs.cornell.edu/home/kleinber/> [cit. 2013-04-25].
- [4] Barabási, Albert-László, *V pavučine sítí*, Paseka, 2005.
- [5] Demel, Jiří, *Grafy a jejich aplikace*, Academia, 2002.
- [6] Palúch, Stanislav, *Algoritmická teória grafov*, Žilinská univerzita, 2008.
- [7] *Network Science*  
Dostupné z: <http://barabasilab.neu.edu/networksciencebook/downloadPDF.html> [cit. 2013-04-25].
- [8] *The structure and function of complex networks*  
Dostupné z: <http://arxiv.org/pdf/cond-mat/0303516.pdf> [cit. 2013-04-25].
- [9] *Collective dynamics of 'small-world' networks*  
Dostupné z: [http://research.yahoo.com/files/w\\_s\\_NATURE\\_0.pdf](http://research.yahoo.com/files/w_s_NATURE_0.pdf) [cit. 2013-04-25].
- [10] Newman, M. E. J., *Networks*, Oxford University Press, 2010.
- [11] *The Poisson Distribution*  
Dostupné z: <http://www.hamilton.ie/ollie/EE304/Poisson.pdf> [cit. 2013-04-25].
- [12] *Grafové algoritmy a komplexní sítě (GAKS)*  
Dostupné z: <http://www.cs.vsb.cz/ochodkova/> [cit. 2013-04-25].
- [13] *Chain links*  
Dostupné z: [http://djjr-courses.wdfiles.com/local-files/soc180:karinthy-chain-links/Karinthy-Chain-Links\\_1929.pdf](http://djjr-courses.wdfiles.com/local-files/soc180:karinthy-chain-links/Karinthy-Chain-Links_1929.pdf) [cit. 2013-04-25].

- 
- [14] *On the properties of small-world network models*  
Dostupné z: <http://www.cpt.univ-mrs.fr/barrat/epjb13.pdf> [cit. 2013-04-25].
- [15] *gnuplot download*  
Dostupné z: <http://www.gnuplot.info/download.html> [cit. 2013-04-25].
- [16] *Correlations and clustering in a scale-free network in Euclidean space*  
Dostupné z: <http://arxiv.org/pdf/cond-mat/0303516.pdf> [cit. 2013-04-25].
- [17] *The structure and function of complex networks*  
Dostupné z: <http://www.ias.ac.in/pramana/v71/p391/fulltext.pdf> [cit. 2013-04-25].
- [18] *Statistical mechanics of complex networks*  
Dostupné z: [http://www.barabasilab.com/pubs/CCNR-ALB\\_Publications/200201-30\\_RevModernPhys-StatisticalMech/200201-30\\_RevModernPhys-StatisticalMech.pdf](http://www.barabasilab.com/pubs/CCNR-ALB_Publications/200201-30_RevModernPhys-StatisticalMech/200201-30_RevModernPhys-StatisticalMech.pdf) [cit. 2013-04-25].
- [19] *Analysis with igraph*  
Dostupné z: <http://igraph.sourceforge.net/igraphbook/igraphbook-random.html> [cit. 2013-04-25].



## A Príloha

